

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2004, 10:20:21 ; Search time 341 Seconds

(without alignments)

4796.353 Million cell updates/sec

Title: M64347

Perfect score: 385

Sequence: 1 GACITCAAGCAAGCTGCTA.....GTGCATGCTGCCAGAGGTG 385

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002as.*

7: Geneseqn2003as.*

8: Geneseqn2003bs.*

9: Geneseqn2003cs.*

10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	385	100.0	3829	6	ABL68525	Ab168525 Kidney ca
2	385	100.0	3829	6	ABQ88193	Abq88193 Human ost
3	385	100.0	3829	7	ADA55418	Ada55418 Human FGF
4	385	100.0	4093	8	ADA02847	Ada02847 Human FGF
5	385	100.0	4093	9	AD572585	Ad572585 Human FGF
6	385	100.0	4093	9	ADC85326	Adc85326 Human Fgf
7	385	100.0	3352	8	ADA02846	Ada02846 Human FGF
8	385	100.0	3352	9	ADB72584	Adb72584 Human FGF
9	385	100.0	3352	9	ADC85325	Adc85325 Mouse Fgf
10	383.4	99.6	3582	6	ADA40522	Ada40522 Human fib
11	291	75.6	2052	5	AA506078	AA506078 Angiotens
12	159.2	41.4	3391	8	ADA02843	Ada02843 Mouse Fgf
13	159.2	41.4	3391	9	ADB72581	Adb72581 Mouse Fgf
14	159.2	41.4	3391	9	ADC85322	Adc85322 Human Map
15	126.2	32.8	530	9	ADB56952	Adb56952 Toxicity-
16	60	15.6	60	6	AAK53820	AAK53820 Murine tr
17	49.4	12.8	238	4	AAK53820	AAK53820 Human spl
18	41.8	10.9	21234	6	AAD32039	Aad32039 Human kin
19	40.2	10.4	1266	5	AAC8112	Aac8112 Human FLE
20	40.2	10.4	1631	5	ABV27515	Abv27515 Human pro
21	40.2	10.4	1631	5	ABV21694	Abv21694 Human pro
22	40	10.4	1895	4	AAH18061	Aah18061 Human CDN
23	39.8	10.3	700	4	AAH93026	Aah93026 Human inf

ALIGNMENTS

RESULT 1

ABL68525
ID ABL68525 standard; DNA; 3829 BP.

XX AC ABL68525;

XX AC ABL68525;

DT 15-MAY-2002 (first entry)

XX DE Kidney cancer related gene sequence SEQ ID NO:6862.

XX DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

XX KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;

XX KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;

XX KW gene; ds.

XX OS Homo sapiens.

XX PN WO200194629-A2.

XX PD 13-DEC-2001.

XX PF 30-MAY-2001; 2001WO-US010838.

XX PF 05-JUN-2000; 2000US-0209473P.

XX PF 05-JUN-2000; 2000US-0209531P.

XX PF 18-SEP-2000; 2000US-0233133P.

XX PF 18-SEP-2000; 2000US-0233617P.

XX PF 20-SEP-2000; 2000US-0234009P.

XX PF 20-SEP-2000; 2000US-0234034P.

XX PF 20-SEP-2000; 2000US-0234052P.

XX PF 22-SEP-2000; 2000US-0234509P.

XX PF 22-SEP-2000; 2000US-0234567P.

XX PF 25-SEP-2000; 2000US-0234923P.

XX PF 25-SEP-2000; 2000US-0235077P.

XX PF 25-SEP-2000; 2000US-0235082P.

XX PF 25-SEP-2000; 2000US-0235134P.

XX PF 25-SEP-2000; 2000US-0235280P.

XX PF 26-SEP-2000; 2000US-0235638P.

XX PF 27-SEP-2000; 2000US-0235711P.

XX PF 27-SEP-2000; 2000US-0235720P.

XX PF 27-SEP-2000; 2000US-0235840P.

XX PF 27-SEP-2000; 2000US-0235863P.

XX PF 28-SEP-2000; 2000US-0236028P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

XX PF 28-SEP-2000; 2000US-0236032P.

PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 28-SEP-2000; 2000US-0236842P.
 PR 28-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 XX
 PA (AVAL-) AVALON PHARM.
 XX
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrisan S;
 PI Soppet DR, Weaver Z;
 XX
 XX WPI; 2002-188264/24.
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.
 XX
 PS Claim 1; SEQ ID NO 6862; 44pp; English.
 XX
 XX The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
 CC tumour
 XX
 SQ Sequence 3829 BP; 776 A; 1100 C; 1163 G; 790 T; 0 U; 0 Other;
 Query Match 100.0%; Score 385; DB 6; Length 3829;
 Best Local Similarity 100.0%; Pred. No. 2.1e-94;
 Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GACTTAAAGCAGCTGGTATTTCTATACAAATCTCTTAATGCTGTGTCCTCCAGGCA 60
 DB 3336 GACTTAAAGCAGCTGGTATTTCTATACAAATCTCTTAATGCTGTGTCCTCCAGGCA 3395
 QY 61 GGGAGACGGTTTCCAGGAGGGCCGCTGTGTCAGGTTCCTGATGTTATTAGATGTT 120
 DB 3396 GGGAGACGGTTTCCAGGAGGGCCGCTGTGTCAGGTTCCTGATGTTATTAGATGTT 3455
 QY 121 ACAAGTTAT 180
 DB 3456 ACAAGTTAT 3515
 QY 181 AGACTTAACACTTCTTACGCAAGCTTCTAGAGTTTATAGCTGTGACCTACCTTTCA 240
 DB 3516 AGACTTAACACTTCTTACGCAAGCTTCTAGAGTTTATAGCTGTGACCTGTACCTTTCA 3575

QY 241 AAGCTTGGAGGAGCGCGTGAATTCAGTTGTTGTTCTGTTCTGTTACTGTTACTGGCCCTGAG 300
 DB 3576 AAGCTTGGAGGAGCGCGTGAATTCAGTTGTTGTTCTGTTCTGTTACTGTTACTGGCCCTGAG 3635
 QY 301 TCTGGCAGCTGTCCCTTGTCTTGCCTGCAGGGCCATGGCTCAGGGTGGTCTCTTCTTTGG 360
 DB 3636 TCTGGCAGCTGTCCCTTGTCTTGCCTGCAGGGCCATGGCTCAGGGTGGTCTCTTCTTTGG 3695
 QY 361 GCCCAGTGCATGTGGCCAGAGTG 385
 DB 3696 GCCCAGTGCATGTGGCCAGAGTG 3720
 RESULT 2
 ABQ88193
 ID ABQ88193 standard; cDNA; 3829 BP.
 XX
 AC ABQ88193;
 XX
 DT 18-SEP-2002 (first entry)
 XX
 DE Human osteoblast differentiation related cDNA SEQ ID NO 100.
 XX
 KW Human; osteoblast; stem cell differentiation; bone tissue deposition;
 KW osteoporosis; osteopathic; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200250301-A2.
 XX
 PD 27-JUN-2002.
 XX
 PF 18-DEC-2001; 2001WO-US048276.
 XX
 PR 18-DEC-2000; 2000US-0255882P.
 PR 24-APR-2001; 2001US-0285691P.
 XX
 PA (GENE-) GENE LOGIC INC.
 PA (PROC) PROCTER & GAMBLE CO.
 XX
 PI Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
 PI Mertz L;
 XX
 DR WPI; 2002-557663/59.
 XX
 PT Use of genes and their expression profiles associated with osteoblast
 PT differentiation for screening modulators bone formation, for diagnosing
 PT or treating e.g. osteoporosis, or as markers for the differentiation
 PT process.
 XX
 PS Claim 1; SEQ ID NO 100; 78pp + Sequence Listing; English.
 XX
 CC The invention relates to genes and their expression profiles are used
 CC for: (a) screening modulators of precursor stem cell differentiation into
 CC osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
 CC deposition of bone tissue, abnormal rate of osteoblast formation or
 CC osteoporosis; or (c) treating or monitoring treatment of the conditions
 CC cited in (b) or monitoring the progression of bone tissue deposition.
 CC Specific conditions include postmenopausal osteoporosis, glucocorticoid
 CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
 CC induced abnormalities in bone formation or bone loss, conditions that
 CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
 CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
 CC or fibrous dysplasia. The present sequence is that of an osteoblast
 CC differentiation associated cDNA marker of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 3829 BP; 776 A; 1100 C; 1163 G; 790 T; 0 U; 0 Other;
 Query Match 100.0%; Score 385; DB 6; Length 3829;
 Best Local Similarity 100.0%; Pred. No. 2.1e-94;

Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAAGCAAGCTGGTATTTTCATACAAATTTCTTAATTGCTGTGTCCTCCAGGCA 60
Db 3336 GACTTCAAAGCAAGCTGGTATTTTCATACAAATTTCTTAATTGCTGTGTCCTCCAGGCA 3395
QY 61 GGGAGACGGTTCACAGAGAGGGCGCGCCCTGTGTCAGGTTCCGAGTTATAGATTT 120
Db 3396 GGGAGACGGTTCACAGAGAGGGCGCGCCCTGTGTCAGGTTCCGAGTTATAGATTT 3455
QY 121 ACAAGTTTATATATATCTATATATATAATTTATTTAGTTTTCACAGATGTTATTTGTTGT 180
Db 3456 ACAAGTTTATATATCTATATATATAATTTATTTAGTTTTCACAGATGTTATTTGTTGT 3515
QY 181 AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGAGCTGCTTTCCTTCA 240
Db 3516 AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGAGCTGCTTTCCTTCA 3575
QY 241 AGCTTGGAGGGAAGCCGGAATTCAGTTGTTCTGTTCTGTAAGTTTACTGGCCCTGAG 300
Db 3576 AGCTTGGAGGGAAGCCGGAATTCAGTTGTTCTGTTCTGTAAGTTTACTGGCCCTGAG 3635
QY 301 TGTGGCAGCTGTCCCTTGCCTTCAGGCGCCATGGCTCAGGTTGCTCTTCTTGGG 360
Db 3636 TGTGGCAGCTGTCCCTTGCCTTCAGGCGCCATGGCTCAGGTTGCTCTTCTTGGG 3695
QY 361 GCCAGTGCATGGTGCCAGAGGTG 385
Db 3696 GCCAGTGCATGGTGCCAGAGGTG 3720

RESULT 3
AAD55418
ID AAD55418 standard; DNA; 3829 BP.
XX AAD55418;
XX 07-AUG-2003 (first entry)
XX Human FGFR-3 DNA #2.
XX Human; antisense; fibroblast growth factor receptor 3; prophylaxis;
KW developmental disorder; hyperproliferative disorder; antisense therapy;
KW FGFR-3; ACH; JTK4; CEK2; cancer; chromosome 4p16.3; gene; ds.
XX Homo sapiens.
XX WO2003023004-A2.
XX 20-MAR-2003.
XX 06-SEP-2002; 2002WO-US028549.
XX 10-SEP-2001; 2001US-00953047.
XX (ISIS-) ISIS PHARM INC.
XX Monia BP, Wyatt JR;
XX WPI; 2003-313244/30.
XX Novel compound targeted to a nucleic acid molecule encoding fibroblast
PT growth factor receptor 3, useful for inhibiting the expression of the
PT receptor and for treating an animal having cancer or developmental
PT disorder.
XX Example 15; Page 89-92; 120pp; English.
XX The invention relates to antisense compounds targetted to a nucleic acid
CC molecule encoding fibroblast growth factor (FGF) receptor 3 (also known
CC as FGFR-3, ACH, JTK4 and CEK2) to inhibit its expression. Antisense
CC compounds of the invention are useful for treating diseases or conditions
CC associated with FGFR-3 such as developmental disorders or

CC hyperproliferative disorders, especially cancer of colorectal, bladder,
CC bone, lung, cervical, breast or skin. They are useful as research
CC reagents, therapeutics, prophylaxis, kits and diagnostics, and as tools
CC in differential and/or combinatorial analyses to elucidate expression
CC patterns of a portion of the genes expressed within cells and tissues.
CC They are also useful in antisense therapy. The present sequence is human
CC FGFR-3 DNA
XX
SQ Sequence 3829 BP; 776 A; 1100 C; 1163 G; 790 T; 0 U; 0 Other;
Query Match 100.0%; Score 385; DB 7; Length 3829;
Best Local Similarity 100.0%; Pred. No. 2.1e-94;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAAGCAAGCTGGTATTTTCATACAAATTTCTTAATTGCTGTGTCCTCCAGGCA 60
Db 3336 GACTTCAAAGCAAGCTGGTATTTTCATACAAATTTCTTAATTGCTGTGTCCTCCAGGCA 3395
QY 61 GGGAGACGGTTCACAGAGAGGGCGCGCCCTGTGTCAGGTTCCGATGTTATAGATTT 120
Db 3396 GGGAGACGGTTCACAGAGAGGGCGCGCCCTGTGTCAGGTTCCGATGTTATAGATTT 3455
QY 121 ACAAGTTTATATATATCTATATATATAATTTATTTAGTTTTCACAGATGTTATTTGTTGT 180
Db 3456 ACAAGTTTATATATCTATATATATAATTTATTTAGTTTTCACAGATGTTATTTGTTGT 3515
QY 181 AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGAGCTGCTTTCCTTCA 240
Db 3516 AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGAGCTGCTTTCCTTCA 3575
QY 241 AAGCTTGGAGGGAAGCCGGAATTCAGTTGTTCTGTTCTGTAAGTTTACTGGCCCTGAG 300
Db 3576 AAGCTTGGAGGGAAGCCGGAATTCAGTTGTTCTGTTCTGTAAGTTTACTGGCCCTGAG 3635
QY 301 TGTGGCAGCTGTCCCTTGCCTTCAGGCGCCATGGCTCAGGTTGCTCTTCTTGGG 360
Db 3636 TGTGGCAGCTGTCCCTTGCCTTCAGGCGCCATGGCTCAGGTTGCTCTTCTTGGG 3695
QY 361 GCCAGTGCATGGTGCCAGAGGTG 385
Db 3696 GCCAGTGCATGGTGCCAGAGGTG 3720
RESULT 4
ADA02847
ID ADA02847 standard; cDNA; 4093 BP.
XX ADA02847;
XX 06-NOV-2003 (first entry)
XX Human FGFR3 carcinoma associated cDNA, SEQ ID NO:1365.
XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ss.
XX Homo sapiens.
XX WO2003057146-A2.
XX 17-JUL-2003.
XX 26-DEC-2002; 2002WO-US041414.
XX 26-DEC-2001; 2001US-00035832.
XX (SAGR-) SAGRES DISCOVERY.
XX Morris DW;
XX WPI; 2003-587068/55.
XX


```
ID ADC85326 standard; DNA; 4093 BP.
XX
AC ADC85326;
XX
DT 01-JAN-2004 (first entry)
XX
XX Human Fgfr3 genomic sequence.
DE
KW Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW secreted; transmembrane; intracellular; ds.
XX
XX Homo sapiens.
XX
XX WO2003045230-A2.
FN
XX
XX 05-JUN-2003.
XX
XX 02-DEC-2002; 2002WO-US038582.
PR
XX
XX 30-NOV-2001; 2001US-00997722.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW, Engelhard EK;
PI
XX
XX WPI; 2003-513603/48.
DR
XX
XX New recombinant nucleic acid comprising a nucleotide sequence of any of
PT the carcinoma-associated (CA) genes, useful for screening for drug
PT candidates for diagnosing or treating carcinomas.
PS
XX
XX Claim 1; SEQ ID NO 112; 983pp; English.
XX
XX The invention relates to a recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the fully defined carcinoma-
CC associated (CA) genes from the 50 tables given in the specification. The
CC CA proteins are secreted, transmembrane or intracellular proteins. The
CC recombinant nucleic acids are useful for screening for drug candidates
CC for diagnosing or treating carcinomas. Sequences given in ADC85315-
CC ADC85514 represent CA genes of the invention.
XX
XX
XX Sequence 4093 BP; 799 A; 1198 C; 1268 G; 828 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 385; DB 9; Length 4093;
Best Local Similarity 100.0%; Pred. No. 2.2e-94;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAGCAAGCTGTGATTTTCATACAAATTTCTTAATGCTGTGTGCCAGGCA 60
DB 3600 GACTTCAAGCAAGCTGTGATTTTCATACAAATTTCTTAATGCTGTGTGCCAGGCA 3659
QY 61 GGGAGACGGTTTCCAGGGAGGGCGGCCCTGTGTGTCAGGTTCCGATGTTATTAGATGTT 120
DB 3660 GGGAGACGGTTTCCAGGGAGGGCGGCCCTGTGTGTCAGGTTCCGATGTTATTAGATGTT 3719
QY 121 ACAAGTTTATATATCTATATATATATATATATATATATATATATATATATATATATAT 180
DB 3720 ACAAGTTTATATATCTATATATATATATATATATATATATATATATATATATATATAT 3779
QY 181 AGACTTTAACTCTTTACGCAATGCTTCTAGAGTTTATAGCTGACGCTACCTTTCA 240
DB 3780 AGACTTTAACTCTTTACGCAATGCTTCTAGAGTTTATAGCTGACGCTACCTTTCA 3839
QY 241 AAGCTTTGGAGGGAAGCGCGTAATTCAGTTGGTTGTTCTGTACTGTACTGGGCCCTGAG 300
DB 3840 AAGCTTTGGAGGGAAGCGCGTAATTCAGTTGGTTGTTCTGTACTGTACTGGGCCCTGAG 3899
QY 301 TCTGGGACGCTGTCCCTTGTCCCTGTCCGTGAGGGCCATGGCTCAGGGTGTCTCTTCTTGGG 360
DB 3900 TCTGGGACGCTGTCCCTTGTCCCTGTCCGTGAGGGCCATGGCTCAGGGTGTCTCTTCTTGGG 3959
QY 361 GCCCAGTGCATGTGGCCAGAGTG 385
```

```
DB 3960 GCCCAGTGCATGTGGCCAGAGTG 3984.
RESULT 7
ADA02846
ID ADA02846 standard; DNA; 33352 BP.
XX
AC ADA02846;
XX
XX 06-NOV-2003 (first entry)
DT
XX
XX Human FGFR3 carcinoma associated gene, SEQ ID NO:1364.
DE
XX
XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
XX prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
XX gene; ds.
XX
XX Homo sapiens.
XX
XX WO2003057146-A2.
FN
XX
XX 17-JUL-2003.
XX
XX 26-DEC-2002; 2002WO-US041414.
XX
XX 26-DEC-2001; 2001US-00035832.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW;
PI
XX
XX WPI; 2003-587068/55.
DR
XX
XX New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
PT
XX
XX Claim 1; SEQ ID NO 1364; 245pp; English.
XX
XX The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 33352 BP; 5993 A; 9766 C; 10292 G; 6073 T; 0 U; 1228 Other;
SQ
Query Match 100.0%; Score 385; DB 8; Length 33352;
Best Local Similarity 100.0%; Pred. No. 4.7e-94;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAGCAAGCTGTGATTTTCATACAAATTTCTTAATGCTGTGTGCCAGGCA 60
DB 24436 GACTTCAAGCAAGCTGTGATTTTCATACAAATTTCTTAATGCTGTGTGCCAGGCA 24495
QY 61 CGGAGACGGTTTCCAGGAGGGCGGCCCTGTGTGAGGTTCCGATGTTATTAGATGTT 120
```


QY 81 GGGCGGCCCTGTGCGAGGTTCCGATGTTATTAGATGTTACAGTTTATATATATCTAT 140
 Db 24845 GGGCTGGCCCTGCACAAAGGTTTCAGATGTTAATAGTTAT-CTGTTACAGTTTATCT 24900
 QY 141 ATATATAATTTATGAGTTTACAAAGATGTTATTTGTTGTAGACTTAAACACTTCTTACGC 200
 Db 24901 ATCTATATTTATGAGTTTACAAAGTTGT-TTTGCTGTAGGCTTAAACACTTCTTACGC 24959
 QY 201 AATGCTTCAGAGTTTATAGCTGGACTGCTACCTTTCAAGCTTGGAGGAGCCGCTG 260
 Db 24960 AGTGTCTTCAGACTTTTATAGCTAGACTGCTACCTTTCAAGCTTGGAGGAGCTGCTG 25019
 QY 261 AATTC-AGTTGTTCTGTTCTGTTACTGTTACTGGGCCCTGAGTCTGGGAGCTGTCCCTTG 319
 Db 25020 AATGCAATTTGTTACTTTTGTACTGTTACTGGGCCCTAGGCTGGTGGCTGTCCCTTG 25079
 QY 320 CTGTC 325
 Db 25080 CCTGTC 25085

RESULT 13

ADB72581
 ID ADB72581 standard; DNA; 33991 BP.

XX ADB72581;
 DT 04-DEC-2003 (first entry)
 XX Mouse Fgfr3 gene.
 XX mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
 KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX Mus sp.

OS WO2003008583-A2.
 PN 30-JAN-2003.

XX 26-DEC-2001; 2001WO-US051291.
 XX 02-MAR-2001; 2001US-00798586.
 FR 23-OCT-2001; 2001US-00004113.
 PR 08-NOV-2001; 2001US-00052482.
 PR 30-NOV-2001; 2001US-00397722.
 PR 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

PA Morris DW, Engelhard EK;

PI WPI; 2003-239337/23.

XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.

XX Claim 1; SEQ ID NO 409; 2304pp; English.

PS The invention relates to a novel recombinant nucleic acid comprising a
 XX nucleotide sequence selected from any of the 660 sequences fully defined
 CC in the specification. A polynucleotide of the invention has cytostatic
 CC activity, and may have a use in gene therapy, or in a vaccine. The
 CC recombinant nucleic acids and polypeptides are useful for treating
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
 CC sarcomas. The present sequence represents a mouse gene of the invention.

XX Sequence 33991 BP; 7578 A; 8539 C; 9524 G; 8350 T; 0 U; 0 Other;

XX Query Match 41.4%; Score 159.2; DB 9; Length 33991;

XX Best Local Similarity 78.8%; Pred. No. 1.2e-32;

XX Matches 241; Conservative 0; Mismatches 58; Indels 7; Gaps 4;

QY 21 TTTTCATACAAATTTCTTAATTTGTTGTCCTCCAGGAGGAGACGGTTTCCAGGAG 80
 Db 24786 TTTTCATACAGATATTTCTTAATTTGTTGTTATTTCCAGGAG-CAGCTGTGCCAGGAA 24844
 QY 81 GGGCGGCCCTGTGCGAGGTTCCGATGTTATTAGATGTTTACAAGTTTATATATATCTAT 140
 Db 24845 GGGCTGGCCCTGCACAAAGGTTTCAGATGTTAATAGTTAT-CTGTTACAGTTTATCT 24900
 QY 141 ATATATAATTTATGAGTTTATAGCTGGACTGCTACCTTTCAAGCTTGGAGGAGCCGCTG 200
 Db 24901 ATCTATATTTATGAGTTTATAGCTGGACTGCTACCTTTCAAGCTTGGAGGAGCTGCTG 24959
 QY 201 AATGCTTCTAGAGTTTATAGCTGGACTGCTACCTTTCAAGCTTGGAGGAGCCGCTG 260
 Db 24960 AGTGTCTTCAGACTTTTATAGCTAGACTGCTACCTTTCAAGCTTGGAGGAGCTGCTG 25019
 QY 261 AATTC-AGTTGTTCTGTTCTGTTACTGTTACTGGGCCCTGAGTCTGGGAGCTGTCCCTTG 319
 Db 25020 AATGCAATTTGTTACTTTTGTACTGTTACTGGGCCCTAGGCTGGTGGCTGTCCCTTG 25079
 QY 320 CTGTC 325
 Db 25080 CCTGTC 25085

RESULT 14

ADB85322
 ID ADC85322 standard; DNA; 33991 BP.

XX ADC85322;

AC ADC85322;

DT 01-JAN-2004 (first entry)

XX Human Map3k8 coding sequence.

DE Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;

XX secreted; transmembrane; intracellular; ds.

OS Homo sapiens.

XX WO2003045230-A2.

XX 05-JUN-2003.

XX 02-DEC-2002; 2002WO-US038582.

XX 30-NOV-2001; 2001US-00997722.

XX (SAGR-) SAGRES DISCOVERY.

PA Morris DW, Engelhard EK;

PI WPI; 2003-513603/48.

XX New recombinant nucleic acid comprising a nucleotide sequence of any of
 PT the carcinoma-associated (CA) genes, useful for screening for drug
 PT candidates for diagnosing or treating carcinomas.

XX Claim 1; SEQ ID NO 108; 983pp; English.

PS The invention relates to a recombinant nucleic acid comprising a
 XX nucleotide sequence selected from any of the fully defined carcinoma-
 CC associated (CA) genes from the 50 tables given in the specification. The
 CC CA proteins are secreted, transmembrane or intracellular proteins. The
 CC recombinant nucleic acids are useful for screening for drug candidates
 CC for diagnosing or treating carcinomas. Sequences given in ADC85215-
 CC ADC85514 represent CA genes of the invention.

XX Sequence 33991 BP; 7578 A; 8540 C; 9523 G; 8350 T; 0 U; 0 Other;

XX Query Match 41.4%; Score 159.2; DB 9; Length 33991;

XX Best Local Similarity 78.8%; Pred. No. 1.2e-32;

XX Matches 241; Conservative 0; Mismatches 58; Indels 7; Gaps 4;

QY 21 TTTTCATACAAATCTTCTTAATCTGTGTCTCCAGCAGGAGAGGGTTTCAGGAG 80
 Db 24786 TTTTCATACAGATTATCTTAATCTGTGTGTCTCCAGCAGGAGAGGGTTTCAGGAG 24844
 QY 81 GGGCGGCGCTGTGTGCGAGGTTCCGATGTTATTAGATGTTACAAAGTTTATATATATATCTAT 140
 Db 24845 GGGCTGGCGCTGCAAGAAGTTTCAGATGTTAATAGTTAT---CTGTACAGTTTATCT 24900
 QY 141 ATATATAATTTATTCAGTTTTCACAGATGTTTGTGTAGACTTAACACTTCTTACGC 200
 Db 24901 ATCTATAATTTATTCAGTTTTCACAGATGTTTGTGTAGACTTAACACTTCTTACGC 24959
 QY 201 AATGCTTCTAGAGTTTATAGCTGGAGTGTCTACCTTTTCAAGCTTGGAGGAGCGGTG 260
 Db 24960 AGTCTTCTAGACTTTTATAGCTGGAGTGTCTACCTTTTCAAGCTTGGAGGAGCGGTG 25019
 QY 261 AATTC-AGTTGGTTCGTTCTGTACTGTACTGGCCCTGTAGTCTGGCAGCTGTCCCTTG 319
 Db 25020 AATGCAATTTTGTACTTTTGTACTGTCTCACTGGCCCTGTAGTCTGGCAGCTGTCCCTTG 25079
 QY 320 CTTCGC 325
 Db 25080 CCTGTC 25085

RESULT 15
 ADB56952/c
 ID ADB56952 standard; DNA; 530 BP.
 XX
 AC ADB56952;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Toxicity-related gene, SEQ ID 1978.
 XX
 KW Toxic; toxin; gene expression profile; hepatotoxicity; liver;
 KW drug screening; toxicity assay; ds.
 XX
 OS Unidentified.
 XX
 PN WO2003064624-A2.
 XX
 PD 07-AUG-2003.
 XX
 PF 31-JAN-2003; 2003WO-US003194.
 XX
 PR 31-JAN-2002; 2002US-00060087.
 PR 15-MAR-2002; 2002US-0364045P.
 PR 15-MAR-2002; 2002US-0364055P.
 PR 30-DEC-2002; 2002US-0436643P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
 XX
 XX WPI; 2003-689530/65.
 XX
 XX Predicting a toxic effect of a compound, useful in identifying toxicity
 PT markers in liver tissues or cells for drug screening and toxicity assays,
 PT comprises preparing gene expression profile of tissue or cells exposed to
 PT the compound.
 XX
 PS Claim 1; SEQ ID NO 1978; 1156pp; English.
 XX
 CC The present invention relates to a method for predicting a toxic effect
 CC of a compound. The method comprises preparing a gene expression profile
 CC of a tissue or cell sample exposed to the compound, and comparing the
 CC gene expression profile to a database comprising SEQ ID 1-4925, where
 CC differential expression of the gene indicates at least one toxic effect.
 CC The method is useful for predicting at least one toxic effect of a
 CC compound, predicting hepatotoxicity or the progression of a toxic effect
 CC of a compound, identifying an agent that modulates the onset or

CC progression of a toxic response, predicting the cellular pathways that a
 CC compound modulates in a cell, and identifying an agent that modulates at
 CC least one activity of a protein. The method and compositions of the
 CC present invention using a database of genes having liver toxin-induced
 CC differential expression, are useful in identifying toxicity markers in
 CC liver tissues or cells for drug screening and toxicity assays. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 530 BP; 173 A; 124 C; 94 G; 139 T; 0 U; 0 Other;

Query Match 32.8%; Score 126.2; DB 9; Length 530;
 Best Local Similarity 69.5%; Pred No. 2.5e-24;
 Matches 230; Conservative 0; Mismatches 93; Indels 8; Gaps 4;
 QY 1 GACTTCAAGCAAGCTGGTA-TTTTCATACAAATCTTCTTAATGCTGTGTCCAGGC 59
 Db 475 GATTTCAAGGTGAACCTGGTATTTTTCATACAGATTATTCTAATGTTATGTTCCAGGC 416
 QY 60 AGGGAGAGCGTTTCCAGGAGGGCGGCTGTGTGCGAGGTTCCGATGTTATTAGATGT 119
 Db 415 AGGGGCTGTGCCCCAGGAGGGGCTGGCCCTGCAAGAGGTTTCAGATGTTAATAGATGT 356
 QY 120 TACAAAGTTTATATATATCTATATATATAATTTATTGAGTTTTTACAGATGTTATTGTTG 179
 Db 355 ---CTGTTACAGTTTATCTATCTATAATTTATTGAGTTTCTCAAGTTGT-TTTCGTG 301
 QY 180 TAGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCCTGGACTGCTACCTTTC 239
 Db 300 TAGGCTTAACACTTCTTATGCAATGCTTCTAGACTTTTATAGCCTGACTGCTACCTTTC 241
 QY 240 AAAGCTTGGAGGAGCGCTGAATTCA--GTTGGTTGTTCTGTACTGTTACTGGGCCT 297
 Db 240 AAAGCTTGGAGAGAGTGTCTGAATGCAATTTTGTGACAGGTTTACTGTTACTGGGCCT 181
 QY 298 GAGCTGGGAGCTGTCCCTTGTGCTGC 328
 Db 180 AGGCTTGGTGGCTGTCCCTTGTGCTGC 150

Search completed: September 22, 2004, 11:20:40
 Job time : 345 secs

; ORGANISM: Mus musculus
US-09-997-722-103

Query Match 41.4%; Score 159.2; DB 12; Length 33991;
Best Local Similarity 78.8%; Pred. No. 9e-32;
Matches 241; Conservative 0; Mismatches 58; Indels 7; Gaps 4;

```
QY 21 TTTTCATCAAAATCTCTTAATTTGCTGTGTCGCCAGGCGAGGAGACGGTTTCCAGGGAG 80
    |||
Db 24786 TTTTCATCAGATTATCTTAATTTGCTGTGTCGCCAGGCGAGGAGACGGTTTCCAGGGAA 24844
    |||

QY 81 GGGCCGGCCCTGTGTCAGGTTCCGATGTTATTAGATTGTACAAGTTTATATATATATAT 140
    |||
Db 24845 GGGCTGGCCCTGCAAGAAGTTTCAGATGTTAATAGTTAT----CTGTTACAAGTTTATCT 24900
    |||

QY 141 ATATATAATTATTGAGTTTATACAGATGTTATTTCTGTAGACTTACACTTCTTACGC 200
    |||
Db 24901 ATCTAATAATTATTGAGTTTATACAGATGTTTCTGTAGACTTAACTTCTCTATGC 24959
    |||

QY 201 AATGCTTCTAGAGTTTATAGCTGTGACGTCTACCTTTCAAAGCTTTGGAGGAAAGCCGTG 260
    |||
Db 24960 AGTGCTTCTAGACTTTTATAGCTAGACTGTACCTTTCAAAGCTTTGGAGACAGTGGTG 25019
    |||

QY 261 AATTC-AGTTGGTCTGTTCTGTACGTGTTACTGGGCGCTGAGTCTGGGCGAGCTGTCCCTTG 319
    |||
Db 25020 AATGCAATTTGTTACTTTTGTACTGTCACTGGGCGCTAGGCTGGGTGGCTGTCCCTTG 25079
    |||

QY 320 CTTGCC 325
    |||
Db 25080 CTTGTC 25085
```

RESULT 15

US-10-085-783A-51002
; Sequence 51002, Application US/10085783A

; Publication No. US20040037841A1

; GENERAL INFORMATION:

; APPLICANT: ChondroGene Inc.

; APPLICANT: Liew, C.C.

; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis

; FILE REFERENCE: 4231/2002

; CURRENT APPLICATION NUMBER: US/10/085,783A

; CURRENT FILING DATE: 2002-02-28

; PRIOR APPLICATION NUMBER: US 60/305,340

; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/275,017

; PRIOR FILING DATE: 2001-03-12

; PRIOR APPLICATION NUMBER: US 60/271,955

; PRIOR FILING DATE: 2001-02-28

; NUMBER OF SEQ ID NOS: 58994

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 51002

; LENGTH: 239

; TYPE: DNA

; ORGANISM: Human

US-10-085-783A-51002

Query Match 40.2%; Score 154.8; DB 13; Length 239;

Best Local Similarity 98.7%; Pred. No. 9.8e-32;

Matches 156; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 220 AGCCTGGACTGTACCTTTCAAAGCTTGGAGGAAAGCCGTGAATTCAGTTGGTTCTCT 279
    |||
Db 1 AGCCTGGACTGTACCTTTCAAAGCTTGGAGGAAAGCCGTGAATTCAGTTGGTTCTCT 60
    |||

QY 280 GTACTGTTACTGGGCGCTGTGAGTCTGGGCGAGCTGTCCCTTGTCTGCTGCGAGGCGCATGGC 339
    |||
Db 61 GTACTGTTACTGGGCGCTGTGAGTCTGGGCGAGCTGTCCCTTGTCTGCTGCGAGGCGCATGGC 120
    |||

QY 340 TCAGGGTGTCTCTCTTCTGGGCGCCAGTGCATGGTGCC 377
    |||
Db 121 TCAGGGTGTCTCTCTTCTGGGCGCCAGTGCATGGTGCC 158
    |||
```

Search completed: September 22, 2004, 12:39:28
Job time : 357 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2004, 10:20:21 ; Search time 1784 Seconds
(without alignments)
9353.738 Million cell updates/sec

Title: M64347

Perfect score: 385

Sequence: 1 GACTTCAAGCAAGCTGTA.....GTGCATGGTGGCCAGAGGTG 385

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	385	100.0	3829	6	AX336353	AX336353 Sequence
2	385	100.0	3829	9	HUMFGRLR	M64347 Human novel
3	385	100.0	4093	6	AX695738	AX695738 Sequence
4	385	100.0	5359	9	HSW805784	BS37709 Homo sapi
5	385	100.0	16976	9	AF487554	AF487554 Homo sapi
6	385	100.0	33352	6	AX695737	AX695737 Sequence
7	385	100.0	189223	9	AC016773	AC016773 Homo sapi
8	291	75.6	2052	6	AX147524	AX147524 Sequence
9	189	49.1	67890	2	AC099840	AC099840 Homo sapi
10	159.2	41.4	33991	6	AX695734	AX695734 Sequence
11	159.2	41.4	201973	2	AC079504	AC079504 Mus muscu
12	159.2	41.4	245130	2	AC079571	AC079571 Mus muscu
13	157.6	40.9	1848	10	MUSFGFR17	L42132 Mus muscu
14	157.6	40.9	4156	10	BC053056	BC053056 Mus muscu
15	157.6	40.9	4158	10	MUSMER3	M81342 BALB/c fibr
16	141	36.6	180035	2	AC125819	AC125819 Rattus no
17	141	36.6	209428	2	AC125948	AC125948 Rattus no
18	131	34.0	222009	2	AC133613	AC133613 Rattus no
19	131	34.0	267692	2	AC135138	AC135138 Rattus no
20	126.2	32.8	175164	2	AC119911	AC119911 Mus muscu
21	74.4	19.3	299974	2	AC079512	AC079512 Mus muscu
22	68.8	17.9	156848	9	AC097710	AC097710 Homo sapi
23	68.8	17.9	160918	2	AC069196	AC069196 Homo sapi
24	49.4	12.8	238	6	AX524355	AX524355 Sequence
25	49.4	12.8	238	6	AX553093	AX553093 Sequence
26	44	11.4	110000	2	PFMAL13_00	AL844509 Plasmodiu
27	44	11.4	110000	2	PFMAL13_01	Continuation (2 of
28	44	11.4	192581	2	PFMAL13P1	AL049180 Plasmodiu
29	43.6	11.3	148926	2	AC102641	AC102641 Mus muscu
30	43.4	11.3	250195	3	AE014831	AE014831 Plasmodiu
31	43.2	11.2	144022	5	EX088594	EX088594 Zebrafish
32	43	11.2	257426	2	EX546500	EX546500 Danio rer
33	42.8	11.1	184798	2	AC139981	AC139981 Rattus no
34	42.8	11.1	234967	2	AC134282	AC134282 Rattus no
35	42.8	11.1	248294	2	AC097690	AC097690 Rattus no
36	42.6	11.1	120465	2	AC137001	AC137001 Oryza sat
37	42.6	11.1	183046	8	AC087599	AC087599 Oryza sat
38	42.6	11.1	286362	2	AC141488	AC141488 Rattus no
39	42.6	11.1	304969	8	AE017120	AE017120 Oryza sat
40	42.4	11.0	110000	2	AC119727_1	Continuation (2 of
41	42.4	11.0	110000	2	AC119727_2	Continuation (3 of
42	42.4	11.0	110000	2	PFMAL7P1_09	Continuation (10 o
43	42.4	11.0	110000	2	PFMAL7P1_10	Continuation (11 o
44	42.4	11.0	168963	2	AC09705	AC09705 Homo sapi
45	42.4	11.0	169102	5	AL929146	AL929146 Zebrafish

ALIGNMENTS

RESULT 1
AX336353
LOCUS AX336353 3829 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 6862 from Patent WO0194629.
ACCESSION AX336353
VERSION AX336353.1 GI:18127072
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigant, S., Soppet, D.R., and Weaver, Z.

Cancer gene determination and therapeutic screening using signature

polyA_site	5290
ORIGIN	
Query Match	100.0%; Score 385; DB 9; Length 5359;
Best Local Similarity	100.0%; Pred.No. 2,8e-87;
Matches 385; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GACTTCAAAGCAAGCTGGTATTTTCATACAAAATCTTCTAAATTCGTGTGTGCCAGCA 60
DB	4702 GACTTCAAAGCAAGCTGGTATTTTCATACAAAATCTTCTAAATTCGTGTGTGCCAGCA 4643
QY	61 GGGAGACGGTTTCCAGGGAGGGCCGGCTGTGCAGGTCGAGTTCGATGTTATTAGATGTT 120
DB	4642 GGGAGACGGTTTCCAGGGAGGGCCGGCTGTGCAGGTCGATGTTATTAGATGTT 4583
QY	121 ACAAGTTTATATATATCTATATATAATAATTTATTAGTGTTCACAGATGTTATTGTTGT 180
DB	4582 ACAAGTTTATATATATCTATATATAATAATTTATTAGTGTTCACAGATGTTATTGTTGT 4523
QY	181 AGACTTTAAACACTTCTTACGCAATGCTCTAGAGTTTATAGCCTGGACCTGCTACCTTCA 240
DB	4522 AGACTTTAAACACTTCTTACGCAATGCTCTAGAGTTTATAGCCTGGACCTGCTACCTTCA 4463
QY	241 AAGCTTGGAGGAAGCGGTGAATTCAGTTGGTTCGTTCTGACTGTTACTGGGCCCTGAG 300
DB	4462 AAGCTTGGAGGAAGCGGTGAATTCAGTTGGTTCGTTCTGACTGTTACTGGGCCCTGAG 4403
QY	301 TCTGGCAGCTGTCCTTGCCTTGCCTGACAGGCCATGCTCAGGTGGTCTCTCTTTGGG 360
DB	4402 TCTGGCAGCTGTCCTTGCCTTGCCTGACAGGCCATGCTCAGGTGGTCTCTCTTTGGG 4343
QY	361 GCCCAGTGCATGGTGCCACAGAGGTG 385
DB	4342 GCCCAGTGCATGGTGCCACAGAGGTG 4318
RESULT 5	
AF487554	16976 bp DNA linear PRI 05-MAY-2002
LOCUS	Homo sapiens fibroblast growth factor receptor 3 (FGFR3) gene,
DEFINITION	partial cds, alternatively spliced.
ACCESSION	AF487554
VERSION	AF487554.1 GI:20452379
KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 16976)
JOURNAL	Li, D.L. and Cox, D.R.
REFERENCE	Fibroblast growth factor receptor 3 (FGFR3) genomic sequence
AUTHORS	Unpublished
TITLE	2 (bases 1 to 16976)
JOURNAL	Li, D.L. and Cox, D.R.
REFERENCE	Direct Submission
AUTHORS	Submitted (26-FEB-2002) Genetics, Stanford University, SUNC M312,
TITLE	300 Pasteur Drive M/C 5120, Stanford, CA 94305-5120, USA
FEATURES	Location/Qualifiers
source	1..16976
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4p16.3"
	<1..>13026
gene	/gene="FGFR3"
mRNA	Join(<5018..5287,5511..5576,7131..7300,7384..7507,
	7599..7789,8678..8828,10094..10284,10588..10733,
	11119..11240,11323..11433,11514..11704,11815..11937,
	12021..12091,12310..12447,12593..12698,12880..>13026)
	/gene="FGFR3"
	/product="fibroblast growth factor receptor 3"
	/note="alternatively spliced"


```

DEFINITION Sequence 1364 from Patent WO03008583.
ACCESSION AX695737
VERSION AX695737.1 GI:29418891
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Morris,D.W. and Engelhard,E.K.
TITLE Novel compositions and methods for cancer
JOURNAL Patent: WO 03008583-A 1364 30-JAN-2003;
Sagres Discovery (US)
FEATURES
    source
        1..33352
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 385; DB 6; Length 33352;
Best Local Similarity 100.0%; Pred. No. 3.3e-87;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAGCAAGCTGGTATTTCATACAAATCTCTTAATGCTGTGTCGCCAGGCA 60
Db 24436 GACTTCAAGCAAGCTGGTATTTCATACAAATCTCTTAATGCTGTGTCGCCAGGCA 24495
QY 61 GGAGACGGTTTCCAGGGAGGGCCGCCCTGTGTCAGGTTCCGATGTTATTAGATGTT 120
Db 24496 GGAGACGGTTTCCAGGGAGGGCCGCCCTGTGTCAGGTTCCGATGTTATTAGATGTT 24555
QY 121 ACAAGTTTATATATCTATATATATAATTAATTAATGATTTTACAAGATGTTTGTGT 180
Db 24556 ACAAGTTTATATATCTATATATATAATTAATTAATGATTTTACAAGATGTTTGTGT 24615
QY 181 AGACTTAAACATCTCTACCAATGCTCTAGAGTTTATAGCTGACGTCACCTTTCA 240
Db 24616 AGACTTAAACATCTCTACCAATGCTCTAGAGTTTATAGCTGACGTCACCTTTCA 24675
QY 241 AGCTTGGAGGGAAGCCGTGAATTCAGTTGGTTGTTCTGTACTGTACTGGCCCTGAG 300
Db 24676 AGCTTGGAGGGAAGCCGTGAATTCAGTTGGTTGTTCTGTACTGTACTGGCCCTGAG 24735
QY 301 TCTGGCAGCTCTCCCTCTGCTGTCAGGCGCATGCTCAGGCGGTCCTCTCTTGGG 360
Db 24736 TCTGGCAGCTCTCCCTCTGCTGTCAGGCGCATGCTCAGGCGGTCCTCTCTTGGG 24795
QY 361 GCCCAGTCATGCTGGCCAGAGGTG 385
Db 24796 GCCCAGTCATGCTGGCCAGAGGTG 24820
RESULT 7
AC016773 189223 bp DNA linear PRI 21-FEB-2002
LOCUS Homo sapiens BAC clone RP11-572017 from 4, complete sequence.
DEFINITION
ACCESSION AC016773
VERSION AC016773.8 GI:18250163
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 189223)
AUTHORS Nguyen,C., Haakenson,W. and Spalding,L.
TITLE The sequence of Homo sapiens BAC clone RP11-572017

```

```

JOURNAL REFERENCE
3 (bases 1 to 189223)
Waterston,R.H.
AUTHORS Direct Submission
JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 189223)
Waterston,R.H.
AUTHORS Direct Submission
JOURNAL Submitted (19-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 189223)
Waterston,R.
AUTHORS Direct Submission
JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 19, 2002 this sequence version replaced gi:17647088.
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0572017
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osceogawa K., Woon,P.Y., Zhao,B., Frengen,E., Tatenio,M., Catanesi,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-572017; actual end is at base position 189223 of RP11-572017.

The sequence from 143509 to 143570 is covered only by a pcr product of clone DNA.

```

FEATURES
    source
        1..189223
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="4"
        /map="4"
        /clone="RP11-572017"
        /clone_lib="RPCI-11"
        misc_feature 1249..2212

```



```

* 15648 15747: gap of 100 bp
* 15748 16499: contig of 752 bp in length
* 15999: gap of 100 bp
* 16000 17345: contig of 746 bp in length
* 17346 17445: gap of 100 bp
* 17446 18126: contig of 681 bp in length
* 18127 18226: gap of 100 bp
* 18227 18979: contig of 753 bp in length
* 18980 19079: gap of 100 bp
* 19080 19806: contig of 727 bp in length
* 19807 19907: gap of 100 bp
* 19907 20649: contig of 743 bp in length
* 20650 20749: gap of 100 bp
* 20750 21476: contig of 727 bp in length
* 21477 21576: gap of 100 bp
* 21577 22298: contig of 722 bp in length
* 22299 23198: gap of 100 bp
* 23199 23110: contig of 712 bp in length
* 23111 23210: gap of 100 bp
* 23211 23939: contig of 729 bp in length
* 23940 24039: gap of 100 bp
* 24040 24779: contig of 740 bp in length
* 24780 24879: gap of 100 bp
* 24880 25622: contig of 743 bp in length
* 25623 25722: gap of 100 bp
* 25723 26458: contig of 736 bp in length
* 26459 26558: gap of 100 bp
* 26559 27255: contig of 697 bp in length
* 27256 27355: gap of 100 bp
* 27356 28101: contig of 746 bp in length
* 28102 28201: gap of 100 bp
* 28202 28940: contig of 739 bp in length
* 28941 29040: gap of 100 bp
* 29041 29796: contig of 756 bp in length
* 29797 29896: gap of 100 bp
* 29897 30631: contig of 735 bp in length
* 30632 30731: gap of 100 bp
* 30732 31435: contig of 704 bp in length
* 31436 31535: gap of 100 bp
* 31536 32243: contig of 708 bp in length
* 32244 32343: gap of 100 bp
* 32344 33071: contig of 728 bp in length
* 33072 33171: gap of 100 bp
* 33172 33897: contig of 726 bp in length
* 33898 33997: gap of 100 bp
* 33998 34725: contig of 728 bp in length
* 34726 34825: gap of 100 bp
* 34826 35569: contig of 744 bp in length
* 35570 35669: gap of 100 bp
* 35670 36411: contig of 742 bp in length
* 36412 36511: gap of 100 bp
* 36512 37231: contig of 720 bp in length
* 37232 37331: gap of 100 bp
* 37332 38076: contig of 745 bp in length
* 38077 38176: gap of 100 bp
* 38177 38897: contig of 721 bp in length
* 38898 38997: gap of 100 bp
* 38998 39715: contig of 718 bp in length
* 39716 39815: gap of 100 bp
* 39816 40539: contig of 724 bp in length
* 40540 40639: gap of 100 bp
* 40640 41341: contig of 702 bp in length
* 41342 41441: gap of 100 bp
* 41442 42168: contig of 727 bp in length
* 42169 42268: gap of 100 bp
* 42269 43012: contig of 744 bp in length
* 43013 43112: gap of 100 bp
* 43113 43827: contig of 715 bp in length
* 43828 43927: gap of 100 bp
* 43928 44651: contig of 724 bp in length
* 44652 44751: gap of 100 bp
* 44752 45464: contig of 713 bp in length
* 45465 45564: gap of 100 bp

```

```

Query Match      49.1%; Score 189; DB 2; Length 67890;
Best Local Similarity 100.0%; Pred. No. 2.6e-37;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTTGTGTGTCGCCAGCA 60
Db 1039 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTTGTGTGTCGCCAGCA 980

QY 61 GGGAGACGGTTTCCAGGAGGGCGCGCCGCTGTGTCAGGTTCCGATCTTATAGATGTT 120
Db 979 GGGAGACGGTTTCCAGGAGGGCGCGCCGCTGTGTCAGGTTCCGATCTTATAGATGTT 920

QY 121 ACAAGTTTATATATATCTATATATATAATTTATTTTACAGATGTTTGTGTTGT 180
Db 919 ACAAGTTTATATATATCTATATATATAATTTATTTTACAGATGTTTGTGTTGT 860

QY 181 AGACTTAAC 189
Db 859 AGACTTAAC 851

```

RESULT 10

```

AX695734      33991 bp      DNA      linear      PAT 31-MAR-2003
Sequence 1361 from Patent WO03008583.
DEFINITION
AX695734
ACCESSION
AX695734.1 GI:29418888
VERSION
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
AUTHORS
Morris, D.W. and Engelhard, E.K.
TITLE
Novel compositions and methods for cancer
JOURNAL
Patent: WO 03008583-A 1361 30-JAN-2003;
Sagres Discovery (US)
FEATURES
Location/Qualifiers
source
1..33991
/organism="Mus musculus"
/mol_type="unassigned DNA"
/db_xref="taxon:10090"
ORIGIN

```

```
Query Match      41.4%; Score 159.2; DB 6; Length 33991;
Best Local Similarity 78.8%; Pred. No. 9.3e-30;
Matches 241; Conservative 0; Mismatches 58; Indels 7; Gaps 4;

QY 21 TTTTCATACAAATCTTCTTAATTCGTGTGTCGCCAGGAGGAGACGGTTTCCAGGGAG 80
    |||||
Db 24786 TTTTCATACAGATTATCTTAATTCGTGTATGTTCCAGGAG-GAGCCTGTGCCAGGAA 24844
    |||||
QY 81 GGCGCGGCGCTGTGTCCAGGTTCCGATGTTATGATGTTACAGTTTATATATATATCTAT 140
    |||||
Db 24845 GGCGTGGCGCTGCAGAGGTTTCAGATGTTAATAGTTAT-CTGTTTAAAGTTTATCT 24900
    |||||
QY 141 ATATATAATTTATGAGTTTTCACAGAGTATTTTGTGTAGACTTAAACACTTCTTAACG 200
    |||||
Db 24901 ATCTATATTTATGAGTTTTCACAGTTGT-TTTGTGTAGGCTTAAACACTTCTTAACG 24959
    |||||
QY 201 AATGCTCTAGAGTTTATAGCTGACCTTCAAGCTTGGAGGAGGAGCGGTG 260
    |||||
Db 24960 AGTGTCTTAGACTTTTATAGCTAGCTGACCTTCAAGCTTGGAGACAGTGTG 25019
    |||||
QY 261 AATTC-AGTGTGTCCTGCTGACTCTTACTGGGCGCTGAGCTGGGCGAGCTGCTCCTTG 319
    |||||
Db 25020 AATGCAATTTTGTACTTTTGTACTGTCTACTGGGCGCTAGGCTGGTGGCTGCTCCTTG 25079
    |||||
QY 320 CTGGCC 325
    |||||
Db 25080 CCGTGC 25085
    |||||

RESULT 11
AC079504/c
LOCUS
DEFINITION Mus musculus clone RP23-256K13, WORKING DRAFT SEQUENCE, 19
ACCESSION AC079504
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 201973)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 201973)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1835429
Center clone name: RPCI-23_256K13
-----
Summary Statistics
Consensus quality: 186841 bases at least Q40
Consensus quality: 195099 bases at least Q30
Consensus quality: 196746 bases at least Q20
Estimated insert size: 208000; agarose-fp estimation
Quality coverage: 6.23 in Q20 bases; sum-of-contigs estimation
Quality coverage: 6.47 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
```

```
* as soon as it is available and the accession number will
* be preserved.
1 1006: contig of 1006 bp in length
1106: gap of unknown length
1107: contig of 1301 bp in length
2407: contig of 1301 bp in length
2408: gap of unknown length
2507: gap of unknown length
4122: contig of 1615 bp in length
4123: gap of unknown length
4223: contig of 1205 bp in length
5427: contig of 1205 bp in length
5527: gap of unknown length
5428: contig of 3093 bp in length
5528: gap of unknown length
8621: gap of unknown length
8721: contig of 1072 bp in length
9792: gap of unknown length
9892: contig of 2097 bp in length
11989: gap of unknown length
12089: gap of unknown length
14322: contig of 2233 bp in length
14323: gap of unknown length
14422: contig of 2056 bp in length
16478: contig of 2056 bp in length
16479: gap of unknown length
16578: gap of unknown length
16579: contig of 3105 bp in length
19683: gap of unknown length
19684: contig of 11549 bp in length
19784: gap of unknown length
31333: gap of unknown length
31334: contig of 9990 bp in length
41323: contig of 12358 bp in length
41422: gap of unknown length
53780: contig of 12358 bp in length
53781: gap of unknown length
53880: gap of unknown length
74766: contig of 20886 bp in length
74767: gap of unknown length
74866: gap of unknown length
93073: contig of 18207 bp in length
93074: gap of unknown length
93173: gap of unknown length
93174: contig of 28180 bp in length
121354: gap of unknown length
121355: contig of 25483 bp in length
146936: gap of unknown length
146937: contig of 25483 bp in length
173036: gap of unknown length
173037: contig of 26265 bp in length
173301: gap of unknown length
173401: gap of unknown length
173302: contig of 28572 bp in length.
173402: contig of 28572 bp in length.

FEATURES
Location/Qualifiers
source
1..201973
/organism="Mus musculus"
/db_type="genomic DNA"
/mb_xref="taxon:10090"
/clone="RP23-256K13"
/clone_lib="RPCI mouse BAC library 23"

ORIGIN
Query Match      41.4%; Score 159.2; DB 2; Length 201973;
Best Local Similarity 78.8%; Pred. No. 1.1e-29;
Matches 241; Conservative 0; Mismatches 58; Indels 7; Gaps 4;

QY 21 TTTTCATACAAATCTTCTTAATTCGTGTGTCGCCAGGAGGAGACGGTTTCCAGGGAG 80
    |||||
Db 149506 TTTTCATACAGATTATCTTAATTCGTGTATGTTCCAGGAG-GAGCCTGTGCCAGGAA 149448
    |||||
QY 81 GGCGCGGCGCTGTGTCCAGGTTCCGATGTTATGATGTTTACAGTTTATATATATCTAT 140
    |||||
Db 149447 GGCGTGGCGCTGCAGAGGTTTCAGATGTTAATAGTTAT-CTGTTTAAAGTTTATCT 149392
    |||||
QY 141 ATATATAATTTATGAGTTTTCACAGAGTATTTTGTGTAGACTTAAACACTTCTTAACG 200
    |||||
Db 149391 ATCTATATTTATGAGTTTTCACAGTTGT-TTTGTGTAGGCTTAAACACTTCTTAACG 149333
    |||||
QY 201 AATGCTCTAGAGTTTATAGCTGACCTTCAAGCTTGGAGGAGGAGCGGTG 260
    |||||
Db 149332 AGTGTCTTAGACTTTTATAGCTAGCTGACCTTCAAGCTTGGAGACAGTGTG 149273
    |||||
QY 261 AATTC-AGTGTGTCCTGCTGACTCTTACTGGGCGCTGAGCTGGGCGAGCTGCTCCTTG 319
    |||||
Db 149272 AATGCAATTTTGTACTTTTGTACTGTCTACTGGGCGCTAGGCTGGTGGCTGCTCCTTG 149213
    |||||
QY 320 CTGGCC 325
    |||||
```


FEATURES		Location/Qualifiers				
source	1..1848	/organism="Mus musculus"				
		/mol_type="genomic DNA"				
		/strain="129"				
		/db_xref="taxon:10090"				
		/clone="35F1"				
		/cell_type="embryonic stem cell"				
		/clone_lib="scos-1"				
	join(L42116.1:207, 211,L42117.1:1, 530,L42118.1:1, 5163, L42119.1:1, 296,L42120.1:1, 1474,L42121.1:1, 102, L42122.1:1, 108,L42123.1:1, 484,L42124.1:1, 364, L42125.1:1, 341,L42126.1:1, 90,L42127.1:1, 93, L42128.1:1, 129,L42129.1:1, 151,L42130.1:1, 244, L42131.1:1, 165,1, .1811)					
		/gene="FGFR3"				
	<1..7	/name="FGFR3"				
intron	/number=18					
	8..182	/gene="FGPR3"				
	/number=18					
exon	183..1811	/gene="FGFR3"				
	/number=19					
	1812..1817					
polyA signal	1831					
polyA_site	/note="The cdna and genomic sequence diverge at this pt."					
ORIGIN						
Query Match 40.9%; Score 157.6; DB 10; Length 1848;						
Best Local Similarity 78.4%; Pred. No. 1.8e-29;						
Matches 240; Conservative 0; Mismatches 59; Indels 7; Gaps 4;						
Qy	21	TTTTTACAAATCTTCTAATTCGTGTGTCCTCCAGGAGGAGACGGTTCCAGGGAG	80			
Db	1397	TTTTTACAGATATATCTAATTCGTATGATTCAGGGAG-GAGCCTGTGCCAGGGAA	1455			
Qy	81	GGCCCGGCCCTGTGTGAGGTTCCGATGTTATAGATGTTACAAAGTTTATATATCTAT	140			
Db	1456	GGGCTGCCCTGCAAGAAGTTTCAGATGTTAATAGTTAT---CTGTTCAGAGTTTATCT	1511			
Qy	141	ATATATATATTTATGATTTTACAGATGTTATTTGTTGAGACTTAACTTCTTACGC	200			
Db	1512	ATCTATATATTTATGAGTTTTCACAGTTGT-TTTGCTGTAGGTTTAACTTCTTATGC	1570			
Qy	201	AATGCTTCTAGAGTTTATAGCTGTGACTCTACCTTTCAAAGCTTGGAGGAGCCGTG	260			
Db	1571	AGTGCTTCTAGACTTTTATAGCTAGACTGCTACCTTTCAAAGCTTGGAGACAGTGGTG	1630			
Qy	261	AATTC-AGTTGTTGTTCTGTTACTGTTACTGGGCCCTGAGTCTGGGAGCTGCCCTTG	319			
Db	1631	AATGCAATTTTGTACTTTTGTACTGTCTACTGGGCCCTAGGCTTGGTGGCTGTCCTTG	1690			
Qy	320	CTTGCC 325				
Db	1691	CCTGTC 1696				
RESULT 14						
BC053056						
LOCUS	BC053056 4156 bp mRNA linear ROD 08-OCT-2003					
DEFINITION	Mus musculus fibroblast growth factor receptor 3, mRNA (cdna clone					
	MGC:62341 IMAGE:5708838), complete cds.					
ACCESSION	BC053056					
VERSION	BC053056.1 GI:31419844					
KEYWORDS	MGC.					
SOURCE	Mus musculus (house mouse)					
ORGANISM	Mus musculus					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	1 (bases 1 to 4156)					
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,					
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,						
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,						
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,						
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,						
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,						
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,						
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,						
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,						
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,						
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,						
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,						
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S.,						
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,						
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,						
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,						
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,						
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.						
Generation and initial analysis of more than 15,000 full-length						
human and mouse cDNA sequences						
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)						
JOURNAL	22388257					
MEDLINE	12477932					
PUBMED	2 (bases 1 to 4156)					
REFERENCE	Strausberg, R.					
AUTHORS	Direct Submission					
JOURNAL	Submitted (02-JUN-2003) National Institutes of Health, Mammalian					
	Gene Collection (MGC), Cancer Genomics Office, National Cancer					
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,					
	USA					
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov					
COMMENT	Contact: MGC help desk					
	Email: cgapbs-remail.nih.gov					
	Tissue Procurement: Dr. Jim Lin, University of Iowa					
	CDNA Library Preparation: M. Bento Soares, University of Iowa					
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)					
	DNA Sequencing by: Sequencing Group at the Stanford Human Genome					
	Center, Stanford University School of Medicine, Stanford, CA 94305					
	Web site: http://www-shgc.stanford.edu					
	Contact: (Dickson, Mark) mcdpaxil@stanford.edu					
	Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,					
	R. M.					
Clone distribution: MGC clone distribution information can be found						
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov						
Series: IRAX Plate: 114 Row: 1 Column: 3						
This clone was selected for full length sequencing because it						
passed the following selection criteria: matched mRNA gi: 6679786.						
FEATURES	Location/Qualifiers					
	source					
gene	1..4156					
	/organism="Mus musculus"					
CDS	/mol_type="mRNA"					
	/strain="C57BL/6"					
gene	/db_xref="taxon:10090"					
	/clone="MGC:62341 IMAGE:5708838"					
CDS	/tissue type="Brain, mouse 15.5 dpc"					
	/clone lib="NIH-BMAP_EXO"					
gene	/lab_host="DH10B"					
	/note="vector: pYX-ASC"					
CDS	1..4156					
	/gene="fgfr3"					
gene	/note="synonym: HBGFR"					
	/db_xref="LocusID:14184"					
CDS	/db_xref="MGI:95524"					
	219..2621					
gene	/codon_start=1					
	/product="fgfr3 protein"					
CDS	/protein_id="AAH53056.1"					
	/db_xref="GI:31419845"					
gene	/db_xref="LocusID:14184"					
	/translation="MVVPACLVFCVAVAGATSEPPGQEVRRVRAAEVPGPEFSQ					
CDS	EQVAFSGDITVELSCHPPGAGPTGTWAKDGTGLVASHRILVGPQRLQVNLASHEDA					
	GVYSCQHLTRVLCHFSVRYTDPAPSSGDDEGDAEDTCAVYWRPERMDKLLAV					
gene	PAANTVRCFAAGNPFPSISLWLNKGFRGEHRIGGKLRHQQNSLVMESVVSFDRG					

NYTCVENKFGSIQTYTLDVLSRPHRPILOAGLPANQTAIILGSDVFEHCKYSDAQ
PHIQWLKHEVNGSKVDPDTPYVTLKTAGANTTDELEVLSHNVTFEDAGEYTCI
ASNPFGSHGSAWLVLPAEELMETDEAGSVYAGVLSYGVVFFLFLVVAAILCRL
RSPKGLGSPVTKHVSFRPLKROVLESNSMNSNTPLVRIARLSGEGPVLANVSE
LELPADPKWELSRTELILGKPLGCGCGOVVMAEAGIDKDRKAKPVTVAVKMKODA
TDKDLSDLVSEMMKMGKXKNIINLLGACTGCGPLVLYVEAAKGNLREFLRARP
PGMDISFACPLPEQLTKDLVSCAVQVARGMEYLASQKCIHRDLAARNVLYTEONV
MKIADFGIARDVHNDLYKKTNGRLPVKMAPEALFDRVYTHQSDVMSFGVLLWEIF
TLGSPGPIGVESLFLKKEGRMDKXPASCTHLYMRECHWAVPSQRTPKQLVE
DLDRILVTSTDEYLDLSVPEQYSPGQDTPSSSSGSDSVFTHDLLPPEPSSNGP
RT"

misc_feature
369..548
/note="IGc2; Region: Immunoglobulin C-2 Type"
/db_xref="CDD:smart00408"

misc_feature
596..938
/note="IG; Region: Immunoglobulin"
/db_xref="CDD:smart00409"

misc_feature
978..1271
/note="IG; Region: Immunoglobulin"
/db_xref="CDD:smart00409"

misc_feature
1614..12444
/note="TyrKc; Region: Tyrosine kinase, catalytic domain"
/db_xref="CDD:smart00219"

ORIGIN

Query Match 40.9%; Score 157.6; DB 10; Length 4156;
Best Local Similarity 78.4%; Pred. No. 2e-29;
Matches 240; Conservative 0; Mismatches 59; Indels 7; Gaps 4;

QY 21 TTTTCATACAAATCTTCTAATCTGTGTGTCGCCAGGAGAGCGTTTCCAGGAG 80
Db 3689 TTTTCATACAGATTATCTAATCTGTATGTTATTCAGGAG-GAGCGTGTGCCAGGAA 3747

QY 81 GGGCGGCGCTGTGCGAGGTCGATGTTTATGATGTTACACTTCTTACGC 200
Db 3748 GGGCTGGCGCTGCAAGAGGTTCAGATGTTAATAGTTAT---CTGTACAAGTTTATCT 3803

QY 141 ATATATAATTTATTCAGTCTTTTACAAGATGTTTGTGATGTTTACACTTCTTACGC 200
Db 3804 ATCTATAATTTATTCAGTCTTTTACAAGTGT-TTGTGTAGCTTAACACTTCTTATGC 3862

QY 201 AATGCTTCAGAGTTTATACCTGGAGCTGCTACCTTTCAAAGCTTGAGAGGAGCGGTG 260
Db 3863 AGTGCTTCAGACTTTTATAGCTTACGCTGCTACCTTTCAAAGCTTGAGAGAGTGGTA 3922

QY 261 AATTC-AGTTGCTGCTTCTGCTACTGTTACTGGGCGCTGAGTGTGGCAGCTGTCCTTG 319
Db 3923 AATGCAATTTTGTACTTTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3982

QY 320 CTGGCC 325
Db 3983 CCTGTC 3988

RESULT 15
MUSMFR3
LOCUS
DEFINITION BALB/c fibroblast growth factor receptor 3 (mFR3) mRNA, complete cds.
ACCESSION M81342 M61881
VERSION M81342.1 GI:199144
KEYWORDS fibroblast growth factor receptor 3; transmembrane protein; tyrosine kinase.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4158)
AUTHORS Ornitz, D.M. and Leder, P.
TITLE Ligand specificity and heparin dependence of fibroblast growth factor receptors 1 and 3
JOURNAL J. Biol. Chem. 267 (23), 16305-16311 (1992)
MEDLINE 9235591

1379594
PURNED
COMMENT Original source text: Mus musculus (strain BALB/c, sub species domesticus) (library: Balb/C brain cDNA library in Lambda ZAP, Stratagene, La Jolla, CA) brain cDNA to mRNA.
FEATURES
Location/Qualifiers
source
1..4158
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/sub_species="domesticus"
/db_xref="taxon:10090"
/tissue_type="brain"
/gene_line
/tissue_lib="Balb/C brain cDNA library in Lambda ZAP, Stratagene, La Jolla, CA"
gene
1..4158
/gene="mFR3"
227..2632
/gene="mFR3"
/codon_start=1
/product="fibroblast growth factor receptor 3"
/protein_id="AAA39535.1"
/db_xref="GI:199145"
/translation="MVVPACVLVFCVAVVAGATSEPPGQEVVRRAAEVGPSPSQ
EQVAFSGDTVELSCHPPGAPGTVMKGTGLVASHRILVGPQRLOVMDKLLAV
GVYSCQHLTRVLCHFSRVYTDAPSSGDDDEDAEDTGAPYWRPMVDKLLAV
PAANTVRPCPAGNPTPSISMLKNGKFRGEHISGIGIKLHQQSILVMSVPSVDRG
NYTCVENKFGSIQTYTLDVLSRPHRPILOAGLPANQTAIILGSDVFEHCKYSDAQ
PHIQWLKHEVNGSKVDPDTPYVTLKTAGANTTDELEVLSHNVTFEDAGEYTCI
AGNSIGFHSASWLVLPAEELMETDEAGSVYAGVLSYGVVFFLFLVVAAILCRL
RSPKGLGSPVTKHVSFRPLKROVLESNSMNSNTPLVRIARLSGEGPVLANVSE
LELPADPKWELSRTELILGKPLGCGCGOVVMAEAGIDKDRKAKPVTVAVKMKODA
TDKDLSDLVSEMMKMGKXKNIINLLGACTGCGPLVLYVEAAKGNLREFLRARP
PGMDISFACPLPEQLTKDLVSCAVQVARGMEYLASQKCIHRDLAARNVLYTEONV
MKIADFGIARDVHNDLYKKTNGRLPVKMAPEALFDRVYTHQSDVMSFGVLLWEIF
TLGSPGPIGVESLFLKKEGRMDKXPASCTHLYMRECHWAVPSQRTPKQLV
EDLDRILVTSTDEYLDLSVPEQYSPGQDTPSSSSGSDSVFTHDLLPPEPSSNG
PRT"

sig_peptide
227..286
/gene="mFR3"

ORIGIN

Query Match 40.9%; Score 157.6; DB 10; Length 4158;
Best Local Similarity 78.4%; Pred. No. 2e-29;
Matches 240; Conservative 0; Mismatches 59; Indels 7; Gaps 4;

QY 21 TTTTCATACAAATCTTCTAATCTGTGTGTCGCCAGGAGAGCGTTTCCAGGAG 80
Db 3702 TTTTCATACAGATTATCTAATCTGTATGTTATTCAGGAG-GAGCGTGTGCCAGGAA 3760

QY 81 GGGCGGCGCTGTGTCAGGTTCCGATGTTTATGATGTTTACAAGTTTATATATATCTAT 140
Db 3761 GGGCTGGCGCTGCAAGAAGGTTCAGATGTTAATAGTTAT---CTGTTGCAAGTTTATCT 3816

QY 141 ATATATAATTTATTCAGTCTTTTACAAGATGTTTGTGATGTTTACACTTCTTACGC 200
Db 3817 ATCTATAATTTATTCAGTCTTTTACAAGTGT-TTGTGTAGCTTAACACTTCTTATGC 3875

QY 201 AATGCTTCAGAGTTTATAGCTGCTGCTACCTTTCAAAGCTTGAGAGGAGCGGTG 260
Db 3876 AGTGCTTCAGACTTTTATAGCTAGACTGCTACCTTTCAAAGCTTGAGAGAGTGGTG 3935

QY 261 AATTC-AGTTGCTGCTTCTGCTACTGTTACTGGGCGCTGAGTGTGGGAGAGTGTCCCTTG 319
Db 3936 AATGCAATTTTGTACTTTTGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3995

QY 320 CTGGCC 325
Db 3996 CCTGTC 4001

Search completed: September 22, 2004, 11:50:38
Job time : 1790 secs

1111

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2004, 10:57:41 ; Search time 2467 Seconds
(without alignments)
4660.292 Million cell updates/sec

Title: M64347

Perfect score: 385

Sequence: 1 GACTTCAAGCAAGCTGGTA.....GTGCAATGGTGGCCAGAGGTG 385

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_prg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	385	100.0	471	10	BF059608
C 2	385	100.0	476	9	AI343936
C 3	385	100.0	479	9	AI078769
C 4	385	100.0	481	10	AW206005

RESULT 1
BF059608/c
LOCUS 7k63d12.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3480238 3',
DEFINITION mRNA sequence.
ACCESSION BF059608
VERSION BF059608.1 GI:10813504
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 471)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmer-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:

ALIGNMENTS

C 5	385	100.0	489	9	AI624729
C 6	385	100.0	492	9	AI080060
C 7	385	100.0	500	10	BG059722
C 8	385	100.0	537	9	AA410992
C 9	385	100.0	544	9	AA411000
C 10	385	100.0	556	12	BQ021570
C 11	385	100.0	575	9	AL043615
C 12	385	100.0	594	14	CA445194
C 13	385	100.0	620	9	AA873489
C 14	385	100.0	649	10	BE466124
C 15	385	100.0	671	9	AI924133
C 16	385	100.0	675	9	AI927305
C 17	385	100.0	680	14	CB851497
C 18	385	100.0	682	9	AA913622
C 19	385	100.0	682	13	BU683104
C 20	385	100.0	683	12	BM670295
C 21	385	100.0	693	13	BQ446455
C 22	385	100.0	700	9	AI521743
C 23	385	100.0	713	12	BM982604
C 24	385	100.0	744	12	BM989149
C 25	385	100.0	744	13	BQ574240
C 26	385	100.0	770	9	AI057095
C 27	383.4	99.6	523	10	AW204106
C 28	381.8	99.2	589	10	AW268519
C 29	381.8	99.2	643	9	AI796742
C 30	381.8	99.2	661	9	AI598047
C 31	381	99.0	544	9	AI332806
C 32	377.6	98.1	553	9	AW009550
C 33	377.6	98.1	721	9	AW074098
C 34	374	97.1	466	10	BF223012
C 35	373.8	97.1	511	9	AA909405
C 36	372.8	96.8	1037	13	BX377171
C 37	371	96.4	601	14	CD674871
C 38	370.4	96.2	691	14	CK005545
C 39	369.8	96.1	475	9	AA928957
C 40	369.8	96.1	504	9	AI199931
C 41	367	95.3	483	9	AI655144
C 42	365.8	95.0	627	10	AW779920
C 43	362.6	94.2	542	9	AA913131
C 44	360.2	93.6	554	9	AI690405
C 45	356.6	92.7	424	9	AA419620

AI624729 ts71b08.x
AI080060 cz32b04.x
BG059722 naf52f12.
AA410992 zv03a11.s
AA411000 zv03b10.s
BQ021570 UI-H-BH1-
AL043615 DKFp434H
CA445194 UI-H-E11-
AA873489 oh77h01.s
BE466124 hy10e11.s
AI924133 wn64a04.x
AI927305 wo89d03.x
CB851497 UI-CF-EN1
AA913622 ol38d10.s
BU683104 UI-CF-EC1
BM670295 UI-E-DW1-
BQ446455 UI-H-EU1-
AI521743 t182d05.x
BM982604 UI-CF-EN1
BM989149 UI-CF-EN0
BQ574240 UI-H-EZ1-
AI057095 oz23f02.x
AW204106 UI-H-B11-
AW268519 xv51b10.x
AI796742 wa14d03.x
AI598047 ts05h09.x
AI332806 qp98d03.x
AW009550 ws84c02.x
AW074098 xb07f11.x
BF223012 7q27d09.x
AA909405 om31g08.s
BX377171 BX377171
CD674871 f816c10.y
CK005545 AGENCOURT
AA928957 o031f10.s
AI199931 qf89e04.x
AI655144 wb54b09.x
AW779920 hn90g11.x
AA913131 om53d10.s
AI690405 tx87c08.x
AA419620 zv04b08.s

info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 454.
Location/Qualifiers
1. 471
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3480238"
/tissue type="pooled germ cell tumors"
/lab host="DH10B"
/note="NCI CGAP GC6"
/note="Vector: p77T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 100.0%; Score 385; DB 10; Length 471;
Best Local Similarity 100.0%; Pred. No. 7e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCCTGCTGTCGCCAGGCA 60
DB 458 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCCTGCTGTCGCCAGGCA 399

QY 61 GGGAGACGGTTTCCAGGAGGGGGCGCCCTGTGTGCGAGTTCGATGTTATTAGATGTT 120
DB 398 GGGAGACGGTTTCCAGGAGGGGGCGCCCTGTGTGCGAGTTCGATGTTATTAGATGTT 339

QY 121 ACAAGTTTATATATATCTATATATATAATTTATTTAGTGTTCACAGATGTTTGTGT 180
DB 338 ACAAGTTTATATATATCTATATATATAATTTATTTAGTGTTCACAGATGTTTGTGT 279

QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTTGGACTGCTACCTTTCA 240
DB 278 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTTGGACTGCTACCTTTCA 219

QY 241 AAGCTTGGAGGAAGCCGCGTAATTCAGTTGGTTCGTTCTGTTCTGTTACTTGGCCCTGAG 300
DB 218 AAGCTTGGAGGAAGCCGCGTAATTCAGTTGGTTCGTTCTGTTCTGTTACTTGGCCCTGAG 159

QY 301 TCTGGGACGCTGCCCTTGTGCTTCAGGCGCCATGCTCAGGTTGGTCTCTTCTTGGG 360
DB 158 TCTGGGACGCTGCCCTTGTGCTTCAGGCGCCATGCTCAGGTTGGTCTCTTCTTGGG 99

QY 361 GCCCAGTGCATGGTGGCCAGAGGTG 385
DB 98 GCCCAGTGCATGGTGGCCAGAGGTG 74

RESULT 2
AI343936/c
LOCUS
DEFINITION
qp42d08.x1 NCI CGAP C08 Homo sapiens cDNA clone IMAGE:1925679 3,
similar to gb:M64347 FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR
(HUMAN); mRNA sequence.

ACCESSION
VERSION
AI343936.1 GI:4081142
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 476)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D.; Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 444.
Location/Qualifiers
1. 476
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1925679"
/tissue type="adenocarcinoma"
/lab host="DH10B"
/clone.lib="NCI CGAP C08"
/note="Organ: colon; Vector: p77T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p77T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 100.0%; Score 385; DB 9; Length 476;
Best Local Similarity 100.0%; Pred. No. 7e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCCTGCTGTCGCCAGGCA 60
DB 467 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCCTGCTGTCGCCAGGCA 408

QY 61 GGGAGACGGTTTCCAGGAGGGGGCGCCCTGTGTGCGAGTTCGATGTTATTAGATGTT 120
DB 407 GGGAGACGGTTTCCAGGAGGGGGCGCCCTGTGTGCGAGTTCGATGTTATTAGATGTT 348

QY 121 ACAAGTTTATATATATCTATATATATAATTTATTTAGTGTTCACAGATGTTTGTGT 180
DB 347 ACAAGTTTATATATATCTATATATATAATTTATTTAGTGTTCACAGATGTTTGTGT 288

QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTTGGACTGCTACCTTTCA 240
DB 287 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTTGGACTGCTACCTTTCA 228

QY 241 AAGCTTGGAGGAAGCCGCGTAATTCAGTTGGTTCGTTCTGTTCTGTTACTTGGCCCTGAG 300
DB 227 AAGCTTGGAGGAAGCCGCGTAATTCAGTTGGTTCGTTCTGTTACTTGGCCCTGAG 168

QY 301 TCTGGGACGCTGCCCTTGTGCTTCAGGCGCCATGCTCAGGTTGGTCTCTTCTTGGG 360
DB 167 TCTGGGACGCTGCCCTTGTGCTTCAGGCGCCATGCTCAGGTTGGTCTCTTCTTGGG 108

QY 361 GCCCAGTGCATGGTGGCCAGAGGTG 385
DB 107 GCCCAGTGCATGGTGGCCAGAGGTG 83

RESULT 3
AI078769/c
LOCUS
DEFINITION
oz34504.x1 Soares, NHMPu, S1 Homo sapiens cDNA clone IMAGE:1677199
3', similar to gb:M64347 FIBROBLAST GROWTH FACTOR RECEPTOR 3
PRECURSOR (HUMAN); mRNA sequence.

ACCESSION
VERSION
AI078769
AI078769.1 GI:3413076

```
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m3 fwd. Et from Amersham
High quality sequence stop: 444.
FEATURES
Location/Qualifiers
1..479
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1677199"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares NhMPu S1"
/notes="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site: 1: Not 1;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NbHPU, and fetal heart NbH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. Clones 260232-265223,
340488-345479, and 484488-489479."
ORIGIN
Query Match 100.0%; Score 385; DB 9; Length 479;
Best Local Similarity 100.0%; Pred. No: 7e-86; Indels 0; Gaps 0;
Matches 385; Conservative 0; Mismatches 0;
QY 1 GACTTCAAAGCAAGCTGGTATTTCATACAAATTCCTTAATTCGTGTGTCCTCCAGGCA 60
Db 474 GACTTCAAAGCAAGCTGGTATTTCATACAAATTCCTTAATTCGTGTGTCCTCCAGGCA 415
QY 61 GGAGACGGTTCCAGGAGGGCGGCGCTGTGTCAGGTTCCGATGTTATAGATGTT 120
Db 414 GGAGACGGTTCCAGGAGGGCGGCGCTGTGTCAGGTTCCGATGTTATAGATGTT 355
QY 121 ACAAGTTTATATATCTATATATATAATTTATTGATTTTACAAGATGTTATTGTTGT 180
Db 354 ACAAGTTTATATATCTATATATATAATTTATTGATTTTACAAGATGTTATTGTTGT 295
QY 181 AGACTTAACTTCTTACGAATGCTTCTAGATTTTATAGCTGACGTGCTACCTTTCA 240
Db 294 AGACTTAACTTCTTACGAATGCTTCTAGATTTTATAGCTGACGTGCTACCTTTCA 235
QY 241 AAGCTTGGAGGGAAGCGTGAATTCAGTTGCTTCTGTTCTGTTACTGTTACTGGCCCTGAG 300
Db 234 AAGCTTGGAGGGAAGCGTGAATTCAGTTGCTTCTGTTCTGTTACTGTTACTGGCCCTGAG 175
QY 301 TCTGGGAGCTGTCCTTGTCTGCCTGCAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 360
Db 174 TCTGGGAGCTGTCCTTGTCTGCCTGCAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 115
QY 361 GCCAGTGCATGGTGGCCAGAGGTG 385
Db 114 GCCAGTGCATGGTGGCCAGAGGTG 90
RESULT 4
```

```
AW206005 481 bp mRNA linear EST 02-DEC-1999
UI-H-B11-afy-f-05-0-UI-s1 NCI CGAP_Sub3 Homo sapiens cDNA clone
IMAGE:2723480 3', mRNA sequence.
ACCESSION AW206005
VERSION AW206005.1 GI:6505479
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html The following repetitive
elements were found in this cDNA sequence: 328-356,
>AT_rich#Low complexity
Seq primer: M13 Forward
POLYA=yes.
Location/Qualifiers
1..481
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2723480"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Sub3"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; The
NCI CGAP Sub3 library is a subtracted library derived from
the NCI CGAP Sub1 library, which is a subtracted library
derived from Bi. Bi constitutes a mixture of 21
normalized or subtracted NCI CGAP libraries:
NCI CGAP Co4, NCI CGAP Pr22, NCI CGAP Pr28, NCI CGAP Co10,
NCI CGAP Co16, NCI CGAP Kid5, NCI CGAP Kid12,
NCI CGAP Kid3, NCI CGAP Kid11, NCI CGAP Lym2,
NCI CGAP Br2, NCI CGAP Co8, NCI CGAP Cui1, NCI CGAP Lei2,
NCI CGAP Brn23, NCI CGAP Lu5, NCI CGAP Lu24,
NCI CGAP Lu19, NCI CGAP GC4, NCI CGAP GC6,
NCI CGAP Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI CGAP Kid3 pool 1 LLAM 3334-3337, 3682-3683,
3793-3803 (IMAGE Clones 1323376-1323911,
1456008-1456775, 1500552-1502855); NCI CGAP Kids pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clones
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI CGAP Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE
Clones 1414920-1417991, 1520904-1522439); NCI CGAP GC4
pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clones 1257096-1258631, 1469064-1470983, LLAM 2457-2459,
1473592-1476743); NCI CGAP Pr22 pool 1
2758-2759, 3062-3068 (IMAGE Clones 985608-986759,
1101192-1101959, 1217928-1220615); NCI CGAP Co10 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clones 1057416-1061255,
1144584-1145351). Subtraction was performed as previously
described (Bonaldo, Lennon & Soares (1996): Normalization
and Subtraction: Two Approaches to Facilitate Gene
Discovery. Genome Research 6, 791-806.
TAG LIB=NCI CGAP_Lu19
TAG_SEQ=GACAGC
```


FEATURES
source Location/Qualifiers
1. .492
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1677007"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/clone_lib="Soares total fetus Nb2HP8_9w"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week TERT fetal material with a Not I - oligo(dT) primer [5', TGTTACCAATCGAGTGGAGCGGCGCTTAATTTTCTTTTCTTTT 3', Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 100.0%; Score 385; DB 9; Length 492;
Best Local Similarity 100.0%; Pred. No. 7e-86; Indels 0; Gaps 0;
Matches 385; Conservative 0; Mismatches 0;

QY 1 GACTTCAAGCAAGCTGGTATTTTCATACAAATTTCTTAATTCGTGTGTGCCAGGCA 60
DB 462 GACTTCAAGCAAGCTGGTATTTTCATACAAATTTCTTAATTCGTGTGTGCCAGGCA 403
QY 61 GGAGACGGTTTCCAGGAGGGCGCGCTGTGTCCAGTTCGGATGTTATAGATTT 120
DB 402 GGAGACGGTTTCCAGGAGGGCGCGCTGTGTCCAGTTCGGATGTTATAGATTT 343
QY 121 ACAAGTTTATATATCTATATATATAATTTATGAGTTTTCACAGATGTTATTTGTTGT 180
DB 342 ACAAGTTTATATATCTATATATATAATTTATGAGTTTTCACAGATGTTATTTGTTGT 283
QY 181 AGACTTAACACTTCTTACGCAATGCTCTAGAGTTTATAGCTGAGCTGCTACCTTCA 240
DB 282 AGACTTAACACTTCTTACGCAATGCTCTAGAGTTTATAGCTGAGCTGCTACCTTCA 223
QY 241 AAGCTTGGAGGAGCGCGTAATTCAGTTGGTTCTGTACTGTACTGGCCCTGAG 300
DB 222 AAGCTTGGAGGAGCGCGTAATTCAGTTGGTTCTGTACTGTACTGGCCCTGAG 163
QY 301 TCTGGGAGCTGTCCTTGTCTGCCTGCAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 360
DB 162 TCTGGGAGCTGTCCTTGTCTGCCTGCAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 103
QY 361 GCCCAGTGCATGTGGCCAGAGGTG 385
DB 102 GCCCAGTGCATGTGGCCAGAGGTG 78

RESULT 7
BG059722/c
LOCUS naf52f12.x1 NCI_CGAP_Brn65 Homo sapiens cDNA clone IMAGE:4147750
DEFINITION 3', mRNA sequence.
ACCESSION BG059722
VERSION BG059722.1 GI:12527487
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 500)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BRGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: ccapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 435.
Location/Qualifiers
1. .500
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4147750"
/tissue_type="glioblastoma without EGFR amplification"
/lab_host="DH10B (TI phage-resistant)"
/clone_lib="NCI_CGAP_Brn65"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.77 kb. Constructed by Life Technologies."

ORIGIN
Query Match 100.0%; Score 385; DB 10; Length 500;
Best Local Similarity 100.0%; Pred. No. 7e-86; Indels 0; Gaps 0;
Matches 385; Conservative 0; Mismatches 0;

QY 1 GACTTCAAGCAAGCTGGTATTTTCATACAAATTTCTTAATTCGTGTGTGCCAGGCA 60
DB 458 GACTTCAAGCAAGCTGGTATTTTCATACAAATTTCTTAATTCGTGTGTGCCAGGCA 399
QY 61 GGGAGACGGTTTCCAGGAGGGCGCGCTGTGTCCAGTTCGGATGTTATAGATTT 120
DB 398 GGGAGACGGTTTCCAGGAGGGCGCGCTGTGTCCAGTTCGGATGTTATAGATTT 339
QY 121 ACAAGTTTATATATCTATATATATAATTTATGAGTTTTCACAGATGTTATTTGTTGT 180
DB 338 ACAAGTTTATATATCTATATATATAATTTATGAGTTTTCACAGATGTTATTTGTTGT 279
QY 181 AGACTTAACACTTCTTACGCAATGCTCTAGAGTTTATAGCTGAGCTGCTACCTTCA 240
DB 278 AGACTTAACACTTCTTACGCAATGCTCTAGAGTTTATAGCTGAGCTGCTACCTTCA 219
QY 241 AAGCTTGGAGGAGCGCGTAATTCAGTTGGTTCTGTACTGTACTGGCCCTGAG 300
DB 218 AAGCTTGGAGGAGCGCGTAATTCAGTTGGTTCTGTACTGTACTGGCCCTGAG 159
QY 301 TCTGGGAGCTGTCCTTGTCTGCCTGCAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 360
DB 158 TCTGGGAGCTGTCCTTGTCTGCCTGCAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 99
QY 361 GCCCAGTGCATGTGGCCAGAGGTG 385
DB 98 GCCCAGTGCATGTGGCCAGAGGTG 74

RESULT 8
AA410992/c
LOCUS AA410992.1
DEFINITION 537 bp mRNA linear EST 18-MAY-1997
similar to gb:M64347 FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION AA410992
VERSION AA410992.1 GI:2070098
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 537)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
Moore,B., Scheellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wyllie,T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997
Unpublished (1997)

TITLE JOURNAL COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: 41ml3 fwd. ET from Amersham

High quality sequence stop: 337.

FEATURES source

Location/Qualifiers
1..537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:5975372"
/db_xref="taxon:9606"
/clone="IMAGE:752540"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares NHMPU S1"
/note="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

ORIGIN

Query Match 100.0%; Score 385; DB 9; Length 537;
Best Local Similarity 100.0%; Pred. No. 7.1e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAAGCAAGCTGTATTTTCATACAAATTTCTTAATGCTGTGTCCAGGCA 60
Db |
466 CACTTCAAAGCAAGCTGTATTTTCATACAAATTTCTTAATGCTGTGTCCAGGCA 407
QY |
61 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGTGAGGTTCCGAGTTATTAGATGTT 120
Db |
406 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGTGAGGTTCCGAGTTATTAGATGTT 347
QY |
121 ACAAGTTTAT 180
Db |
346 ACAAGTTTAT 287
QY |
181 AGACTTTACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGACTGCTACCTTTCA 240
Db |
286 AGACTTTACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGACTGCTACCTTTCA 227
QY |
241 AAGCTTGGAGGAAGCGGTGAATTCAGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 300
Db |
226 AAGCTTGGAGGAAGCGGTGAATTCAGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 167
QY |
301 TCTGGGACGTGTCCTTGTCTGCTGCTGAGGGCCATGGCTCAGGGTGGTCTTCTTGGG 360
Db |
166 TCTGGGACGTGTCCTTGTCTGCTGCTGAGGGCCATGGCTCAGGGTGGTCTTCTTGGG 107
QY |
361 GCCCAGTGCATGGTGGCCAGAGGTG 385
Db |
106 GCCCAGTGCATGGTGGCCAGAGGTG 82

RESULT 9

AA411000/c
LOCUS
DEFINITION

AA411000 544 bp mRNA linear EST 18-MAY-1997
zv03b10.s1 Soares NHMPU S1 Homo sapiens cDNA clone IMAGE:752539 3'
similar to gb:M64347 FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR
(HUMAN); mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AA411000.1 GI:2070106
EST.
Homo sapiens (human)

REFERENCE
AUTHORS

1 (bases 1 to 544)
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
Moore,B., Scheellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wyllie,T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997
Unpublished (1997)

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: 41ml3 fwd. ET from Amersham
High quality sequence stop: 457.

Location/Qualifiers
1..544

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:5975371"
/db_xref="taxon:9606"
/clone="IMAGE:752539"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares NHMPU S1"
/note="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

FEATURES
source

Location/Qualifiers
1..544
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:5975371"
/db_xref="taxon:9606"
/clone="IMAGE:752539"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares NHMPU S1"
/note="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

ORIGIN

Query Match 100.0%; Score 385; DB 9; Length 544;
Best Local Similarity 100.0%; Pred. No. 7.1e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAAGCAAGCTGTATTTTCATACAAATTTCTTAATGCTGTGTCCAGGCA 60
Db |
466 CACTTCAAAGCAAGCTGTATTTTCATACAAATTTCTTAATGCTGTGTCCAGGCA 407
QY |
61 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGTGAGGTTCCGAGTTATTAGATGTT 120
Db |
406 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGTGAGGTTCCGAGTTATTAGATGTT 347
QY |
121 ACAAGTTTAT 180
Db |
346 ACAAGTTTAT 287
QY |
181 AGACTTTACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGACTGCTACCTTTCA 240
Db |
286 AGACTTTACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGACTGCTACCTTTCA 227

ORIGIN

Query Match 100.0%; Score 385; DB 9; Length 544;
Best Local Similarity 100.0%; Pred. No. 7.1e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAAGCAAGCTGTATTTTCATACAAATTTCTTAATGCTGTGTCCAGGCA 60
Db |
466 CACTTCAAAGCAAGCTGTATTTTCATACAAATTTCTTAATGCTGTGTCCAGGCA 407
QY |
61 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGTGAGGTTCCGAGTTATTAGATGTT 120
Db |
406 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGTGAGGTTCCGAGTTATTAGATGTT 347
QY |
121 ACAAGTTTAT 180
Db |
346 ACAAGTTTAT 287
QY |
181 AGACTTTACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGACTGCTACCTTTCA 240
Db |
286 AGACTTTACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGACTGCTACCTTTCA 227

```

QY 241 AAGCTTGGAGGAGCGTGAATTCAGTTGGTTGCTGTTCTGTTACTGTTACTGGGCCCTGAG 300
Db 226 AAGCTTGGAGGAGGAGCGTGAATTCAGTTGGTTGCTGTTCTGTTACTGTTACTGGGCCCTGAG 167
QY 301 TCTGGGAGGTGTCCTTGTGCTGCTGAGGGCCATGGGTCAGGGTGGTCTCTCTTTGGG 360
Db 166 TCTGGGAGGTGTCCTTGTGCTGCTGAGGGCCATGGGTCAGGGTGGTCTCTCTTTGGG 107
QY 361 GCCAGTGCATGTGGCCAGAGTG 385
Db 106 GCCAGTGCATGTGGCCAGAGTG 82

RESULT 10
LOCUS BQ021570/c
DEFINITION UI-H-DH1-axg-j-15-0-UI.sl NCI_CGAP_DH1 Homo sapiens cDNA clone
IMAGE:5828462 3', mRNA sequence.
ACCESSION BQ021570
VERSION BQ021570.1 GI:19756848
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 556)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
The following repetitive elements were found in this cDNA
sequence: 323-351, >AT rich#low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES             Location/Qualifiers
     source           1..556
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5828462"
                     /tissue_type="Metastatic Chondrosarcoma"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_DH1"
                     /notes="Organ: Lung; Vector: p7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI CGAP DH1 is a normalized cDNA library containing the
following tissue(s): VS-8 Cell line from Metastatic
Chondrosarcoma in Lung. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into p7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is AGATCATTCG.
TAG TISSUE=lung
TAG LIB=UI-H-DH1
TAG_SEQ=AGATCATTCG"

ORIGIN
Query Match      100.0%; Score 385; DB 12; Length 556;
Best Local Similarity 100.0%; Pred. No. 7.2e-86;

```

```

Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCRAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCGTGTGTCACAGGCA 60
Db 476 GACTTCRAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCGTGTGTCACAGGCA 417
QY 61 GGGAGACGGTTTCCAGGAGGGGGCCGCTGTGTGACAGTTCGAGTGTATTAAGATT 120
Db 416 GGGAGACGGTTTCCAGGAGGGGGCCGCTGTGTGACAGTTCGAGTGTATTAAGATT 357
QY 121 ACAAGTTTATATATATCTATATATATATATATATATATATATATATATATATATAT 180
Db 356 ACAAGTTTATATATATCTATATATATATATATATATATATATATATATATATATAT 297
QY 181 AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTTATAGCTGGAGTGTACCTTTCA 240
Db 296 AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTTATAGCTGGAGTGTACCTTTCA 237
QY 241 AAGCTTGGAGGAGGAGCGTGAATTCAGTTGGTTGCTGTTACTGTTACTGGGCCCTGAG 300
Db 236 AAGCTTGGAGGAGGAGCGTGAATTCAGTTGGTTGCTGTTACTGTTACTGGGCCCTGAG 177
QY 301 TCTGGGAGCTGTCCCTTGTCTTGCCTGCAGGCCCATGGCTCAGGTTGGTCTTCTTGGG 360
Db 176 TCTGGGAGCTGTCCCTTGTCTTGCCTGCAGGCCCATGGCTCAGGTTGGTCTTCTTGGG 117
QY 361 GCCAGTGCATGTGGCCAGAGTG 385
Db 116 GCCAGTGCATGTGGCCAGAGTG 92

RESULT 11
LOCUS AL043615/c
DEFINITION DKFZp434H1627.s1 434 (synonym: htes3) Homo sapiens cDNA clone
IMAGE:5828462 3', mRNA sequence.
ACCESSION AL043615
VERSION AL043615.1 GI:5423002
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 575)
AUTHORS Bloecker,H., Boecker,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and
Wiemann,S.
TITLE EST (Bloecker, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: MIPS
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 3' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
x1 sequence also available.
This clone (DKFZp434H1627) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source           1..575
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434H1627"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="434 (synonym: htes3)"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

ORIGIN

```


Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 858 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham
High quality sequence stop: 443.
Location/Qualifiers

FEATURES

1. .620

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:1473073"

/lab_host="DH10B"

/clone_lib="NCI CGAP Kid3"

/notes="Organ: kidney; Vector: pTT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTT3 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 385; DB 9; Length 620;

Best Local Similarity 100.0%; Pred. No. 7.3e-86;

Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAGCAAGCTGGTATTTCATACAAATCTTCTAATTCGTGTGCCAGGCA 60

Db 466 GACTTCAAGCAAGCTGGTATTTCATACAAATCTTCTAATTCGTGTGCCAGGCA 407

QY 61 GGGAGACGGTTCCAGGAGGGGGCCCGCCCTGTGTCAGGTCGGATGTTAGATGT 120

Db 406 GGGAGACGGTTCCAGGAGGGGGCCCGCCCTGTGTCAGGTCGGATGTTAGATGT 347

QY 121 ACAAGTTTATATATATCTATATATATAATTAATTTAGTTTTCACAGATGTTTGT 180

Db 346 ACAAGTTTATATATCTATATATATAATTAATTTAGTTTTCACAGATGTTTGT 287

QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGACTGCTACCTTCA 240

Db 286 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGACTGCTACCTTCA 227

QY 241 AAGCTTGGAGGAAGCCGTGAATTCAGTTGGTTGTTCTGTACTGTTACTGGCCCTGAG 300

Db 226 AAGCTTGGAGGAAGCCGTGAATTCAGTTGGTTGTTCTGTACTGTTACTGGCCCTGAG 167

QY 301 TCTGGGAGCTGTCCCTTGTCTGCTCGAGGGCAATGGCTCAGGGTGGTCTCTCTTGGG 360

Db 166 TCTGGGAGCTGTCCCTTGTCTGCTCGAGGGCAATGGCTCAGGGTGGTCTCTCTTGGG 107

QY 361 GCCCAGTGCATGTGGCCAGAGGTG 385

Db 106 GCCCAGTGCATGTGGCCAGAGGTG 82

RESULT 14

BE466124/c

LOCUS

DEFINITION BE466124 649 bp mRNA linear EST 27-JUL-2000
hy10e11.k1 NCI CGAP G66 Homo sapiens cDNA clone IMAGE:3196940 3',
similar to gb:M64347 FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR
(HUMAN); mRNA sequence.

ACCESSION BE466124

VERSION BE466124.1

KEYWORDS EST

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 649)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov

Seq primer: -40UP from Gibco

High quality sequence stop: 454.

Location/Qualifiers

1. .649

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3196940"

/tissue_type="pooled germ cell tumors"

/lab_host="DH10B"

/clone_lib="NCI CGAP G66"

/note="Vector: pTT3D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA

from the normalized library NCI-CGAP G64 was prepared, and

ss circles were made in vitro. Following HAP purification,

this DNA was used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from a pool

of 5,000 clones made from the same library (clones

1257096-1258631, 1469064-1470983, and 1475592-1476743).

Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 385; DB 10; Length 649;

Best Local Similarity 100.0%; Pred. No. 7.4e-86;

Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAGCAAGCTGGTATTTCATACAAATCTTCTAATTCGTGTGCCAGGCA 60

Db 458 GACTTCAAGCAAGCTGGTATTTCATACAAATCTTCTAATTCGTGTGCCAGGCA 399

QY 61 GGGAGACGGTTTCCAGGAGGGGGCCCGCCCTGTGTCAGGTTCCGATGTTAGATGT 120

Db 398 GGGAGACGGTTTCCAGGAGGGGGCCCGCCCTGTGTCAGGTTCCGATGTTAGATGT 339

QY 121 ACAAGTTTATATATCTATATATAATTTAGTTTTCACAGATGTTTGTGT 180

Db 338 ACAAGTTTATATATCTATATATAATTTAGTTTTCACAGATGTTTGTGT 279

QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGACTGCTACCTTCA 240

Db 278 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGACTGCTACCTTCA 219

QY 241 AAGCTTGGAGGAAGCCGTGAATTCAGTTGGTTGTTCTGTACTGTTACTGGCCCTGAG 300

Db 218 AAGCTTGGAGGAAGCCGTGAATTCAGTTGGTTGTTCTGTACTGTTACTGGCCCTGAG 159

QY 301 TCTGGGAGCTGTCCCTTGTCTGCTCGAGGGCAATGGCTCAGGGTGGTCTCTCTTGGG 360

Db 158 TCTGGGAGCTGTCCCTTGTCTGCTCGAGGGCAATGGCTCAGGGTGGTCTCTCTTGGG 99

QY 361 GCCCAGTGCATGTGGCCAGAGGTG 385

Db 98 GCCCAGTGCATGTGGCCAGAGGTG 74

RESULT 15	671 bp	mRNA	linear	EST 08-MAR-2000
AI924133/c				
LOCUS				
DEFINITION	wn54a04.x1 NCI CGAP Lul9 Homo sapiens cDNA clone IMAGE:2450190 3', similar to gb:M64347 FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR (HUMAN);, mRNA sequence.			
ACCESSION	AI924133			
VERSION	AI924133.1	GI:5660097		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 671)			
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index			
JOURNAL	Unpublished (1997)			
COMMENT	Contact: Robert Strausberg, Ph.D. ^a Email: cgapbs- rc@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmer-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Insert Length: 593 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 465.			

ORIGIN

Query Match	100.0%;	Score 385;	DB 9;	Length 671;
Best Local Similarity	100.0%;	Pred. No. 7.4e-86;		
Matches 385;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
1	GACITCAAAGCAAGCTGGTATTTTTCATACAAATTCITCTAATTCGTGTGTCACAGCA	60		
	1			
459	GACITCAAAGCAAGCTGGTATTTTTCATACAAATTCITCTAATTCGTGTGTCACAGCA	400		
	1			
61	GGGAGACGGTTTCCAGGAGGGCGCGCCCTGTGTCACAGTTCGATCTTATTAGATGTT	120		
	1			
399	GGGAGACGGTTTCCAGGAGGGCGCGCCCTGTGTCACAGTTCGATCTTATTAGATGTT	340		
	1			
121	ACAAAGTTTATATATCTATATATAATTTATTCAGTTTTTACAAAGATGATTTGTTGT	180		
	1			
339	ACAAAGTTTATATATCTATATATAATTTATTCAGTTTTTACAAAGATGATTTGTTGT	280		
	1			
181	AGACTTAACACTCTCTTACGCAATGCTCTTAGAGTTTTTATAGCTGGACTGCTACCTTTCA	240		
	1			
279	AGACTTAACACTCTCTTACGCAATGCTCTTAGAGTTTTTATAGCTGGACTGCTACCTTTCA	220		
	1			

241	QY	AAAGCTTGAGAGGAGCCGCGAAATCAGTTGTGTCGTTCTGTACTCTTACTAGGCCCCGAG	300
219	Db	AAAGCTTGAGAGGAGCCGCGAAATCAGTTGTGTCGTTCTGTACTCTTACTAGGCCCCGAG	160
301	QY	TCTGGGACAGTGCCTTCGTTGCTGCTCAGGGCCATGCTCAGGGTGTCTCTTCTTGGG	360
159	Db	TCTGGGACAGTGCCTTCGTTGCTGCTCAGGGCCATGCTCAGGGTGTCTCTTCTTGGG	100
361	QY	GCCCAAGTGCATGTGTGGCCAGAGGTG	385
99	Db	GCCCAAGTGCATGTGTGGCCAGAGGTG	75

Search completed: September 22, 2004, 12:31:55
Job time : 2471 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2004, 11:06:21; Search time 76 Seconds
(without alignments)
2811.264 Million cell updates/sec

Title: M64347

Perfect score: 385

Sequence: 1 GACTTCAAGCAAGCTGTTA.....GTGCATGGTGGCCAGAGGTG 385

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents NA.*

1: /cgn2_6/prodata/2/ina/5A COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B COMB.seq.*
5: /cgn2_6/prodata/2/ina/PCTUS COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfileseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41.8	10.9	21234	4	US-09-810-671-3
2	41.8	10.9	21234	4	US-10-109-854-3
3	37.4	9.7	3323	3	US-08-860-635A-20
4	37.4	9.7	3323	4	US-09-281-476-20
5	36.6	9.5	4839	4	US-09-353-332-1
6	35.8	9.3	6038	3	US-09-305-639-4
7	35.8	9.3	6038	4	US-09-525-1608-2
8	35.8	9.3	7622	3	US-09-305-639-1
9	35.8	9.3	7622	4	US-09-525-1608-1
10	35.8	9.3	8791	5	PCT-US96-01735-5
11	35.8	9.3	9610	4	US-09-566-921-45
12	35.6	9.2	3627	1	US-08-104-072B-6
13	35.6	9.2	3627	1	US-08-351-413-7
14	35.6	9.2	3627	2	US-09-025-583-7
15	35.2	9.1	6368	4	US-10-204-708-67
16	35.2	9.1	640681	4	US-09-790-988-1
17	35	9.1	84495	4	US-09-797-908-3
18	34.4	8.9	1522	4	US-09-620-312D-96
19	34.4	8.9	19233	4	US-10-204-708-45
20	34.2	8.9	2614	4	US-09-004-056-1
21	34	8.8	1245	4	US-09-107-532A-2717
22	33.8	8.8	8607	4	US-10-204-708-71
23	33.6	8.7	53332	4	US-09-801-861-3
24	33.4	8.7	8537	4	US-10-204-708-41
25	33.2	8.6	870	4	US-08-956-171E-879
26	33.2	8.6	3416	2	US-08-451-822A-15
27	33.2	8.6	3416	4	US-08-323-430-15

```

28 33.2 8.6 5152 4 US-10-204-708-73 Sequence 73, Appl
29 33.2 8.6 5585 1 US-08-030-096-3 Sequence 3, Appl
c 30 33 8.6 5095 1 US-08-092-817-3 Sequence 3, Appl
c 31 33 8.6 5095 4 US-08-485-128-3 Sequence 3, Appl
c 32 33 8.6 63000 4 US-09-780-172-18 Sequence 18, Appl
c 33 32.8 8.5 483 4 US-09-404-879A-114 Sequence 114, Appl
c 34 32.8 8.5 483 4 US-09-338-933-114 Sequence 114, Appl
c 35 32.8 8.5 483 4 US-09-215-681-114 Sequence 114, Appl
c 36 32.8 8.5 483 4 US-09-216-003A-114 Sequence 114, Appl
c 37 32.8 8.5 1368 3 US-08-874-563-5 Sequence 5, Appl
c 38 32.8 8.5 1368 3 US-08-577-483-14 Sequence 14, Appl
c 39 32.8 8.5 5983 4 US-09-799-875-4 Sequence 4, Appl
c 40 32.8 8.5 11131 4 US-10-204-708-28 Sequence 28, Appl
c 41 32.8 8.5 11485 4 US-09-410-464-9 Sequence 9, Appl
c 42 32.4 8.4 1818 4 US-09-357-206A-6 Sequence 6, Appl
c 43 32.4 8.4 2876 3 US-08-840-204-1 Sequence 1, Appl
c 44 32.4 8.4 2876 3 US-08-840-204-10 Sequence 10, Appl
c 45 32.4 8.4 2876 4 US-09-324-494A-1 Sequence 1, Appl

```

ALIGNMENTS

RESULT 1

```

US-09-810-671-3
; Sequence 3, Application US/09810671
; Patent No. 6455291
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CLO00758
; CURRENT APPLICATION NUMBER: US/09/810,671
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 21234
; TYPE: DNA
; ORGANISM: Human
US-09-810-671-3

```

```

Query Match 10.9%; Score 41.8; DB 4; Length 21234;
Best Local Similarity 61.5%; Pred. No. 0.011;
Matches 67; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 121 ACAAGTTATATATATCTATATATATATATATATATATGAGCTTTTACAGATGATTTTGTGT 180
Db 12135 ACTAATTATATATATATATATATATATATATATATATTTAAAAATTTTATTTTGTAT 12194

Qy 181 AGACTTAACACTTCTTAGCAATGCTTCTAGAGTTTATAGCCTGGACT 229
Db 12195 TTGTTTATTTTATTTTGTAGAGGAGTTTGTCTTTATGCCCAGGCT 12243

```

RESULT 2

```

US-10-109-854-3
; Sequence 3, Application US/10109854
; Patent No. 6630337
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CLO00758DIV
; CURRENT APPLICATION NUMBER: US/10/109,854
; CURRENT FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 60/227,470
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/810,671
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 5

```



```

RESULT 7
US-09-525-160B-2/c
; Sequence 2, Application US/09525160B
; Patent No. 6569681
; GENERAL INFORMATION:
; APPLICANT: Ivanov, Evgenii
; TITLE OF INVENTION: METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION
; FILE REFERENCE: 10278/016001
; CURRENT APPLICATION NUMBER: US/09/525.160B
; CURRENT FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6038
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-525-160B-2

Query Match          9.3%; Score 35.8; DB 4; Length 6038;
Best Local Similarity 61.1%; Pred.No.0.44;
Matches 58; Conservative 0; Mismatches 37; Indels 0; Gaps 0

QY 101 TTC CGATGTTATTAGATGTTACAGTTTATATATCTATATATCTATATATAATTTATTCGATTT 160
Db 503 TTCCTTTATGTTTGGTTTATATATATATATATATATATATATATATATATATATATATGTTT 444

QY 161 TTACAAGATGATTGTTGTAGACTTAAACACTTCT 195
Db 443 ATATATTATATAATAATATATATATTTATATATTAT 409

RESULT 8
US-09-305-639-1/c
; Sequence 1, Application US/09305639
; Patent No. 6200778
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Selden, Richard F.
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/016001
; CURRENT APPLICATION NUMBER: US/09/305.639
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,663
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 7622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-305-639-1

Query Match          9.3%; Score 35.8; DB 3; Length 7622;
Best Local Similarity 61.1%; Pred.No.0.5;
Matches 58; Conservative 0; Mismatches 37; Indels 0; Gaps 0

QY 101 TTC CGATGTTATTAGATGTTACAGTTTATATATCTATATATCTATATATAATTTATTCGATTT 160
Db 503 TTCCTTTATGTTTGGTTTATATATATATATATATATATATATATATATATATATATATGTTT 444

QY 161 TTACAAGATGATTGTTGTAGACTTAAACACTTCT 195
Db 443 ATATATTATATAATAATATATATATTTATATATTAT 409

RESULT 9
US-09-525-160B-1/c
; Sequence 1, Application US/09525160B
; Patent No. 6569681
; GENERAL INFORMATION:
; APPLICANT: Ivanov, Evgenii

```


APPLICANT: Scheirlinck, Trees
APPLICANT: Komari, Tosihiko
TITLE OF INVENTION: Stamen-specific Promoters from Rice
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 No. 5639948west Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/104,072B
FILING DATE: 05-AUG-1993
CLASSIFICATION: 800
PRIOR APPLICATION DATA: WO 9200272
FILING DATE: 06-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 91403352.7
FILING DATE: 10-DEC-1991
PRIOR APPLICATION DATA: EP 91402590.3
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 91400318.1
FILING DATE: 08-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kowalchuk, Katherine M.
REGISTRATION NUMBER: 36,848
REFERENCE/DOCKET NUMBER: 8076.93USWO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3627 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: promoter
LOCATION: 1..2845
OTHER INFORMATION: /function= "sequence comprising
OTHER INFORMATION: anther-specific PT72 promoter"
FEATURE:
NAME/KEY: TATA signal
LOCATION: 2733..2739
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2765
OTHER INFORMATION: /product= "transcription
OTHER INFORMATION: initiation"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2846
OTHER INFORMATION: /product= "ATG start translation of
OTHER INFORMATION: T72 gene"
US-08-104-072B-6
Query Match 9.2%; Score 35.6; DB 1; Length 3627;
Best Local Similarity 47.0%; Pred. No. 0.39;
Matches 110; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 1 GACTTCAAGCAAGCTGGTATTTCATACAAATCTCTAAATTCCTGCTGTGTCCTCCAGGCA 60

Db 1188 GAGAGCAATGCAAGTTCACAAACAAACAGGTCCTTCCAAACGGGTGGTTTCATTTCCA 1247
QY 61 GGGAGACGGTTTCCAGGAGGGGCGCCCTGTGTGCGAGTTCGAGTGTATTAGATGTT 120
Db 1248 CAGAACAGGATAGCAACCCAGAGACAAACCGTTCAACAATATATATATATATATAT 1307
QY 121 ACAAGTTTATATATATCTATATATATATATATATATATATATATATATATATATAT 180
Db 1308 AT 1367
QY 181 AGACTTAACACTTCTACGCAATGCTCTAGAGTTTATAGCTGGACTGCTAC 234
Db 1368 ATACGGAACCTTAAACACATGAATCTAAACATTTTCAACCAATCAGACTAC 1421
RESULT 13
US-08-351-413-7
; Sequence 7, Application US/08351413
; Patent No. 5750867
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Leemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,413
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/899,072
; FILING DATE: 12-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/970,849
; FILING DATE: 03-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-102PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3627 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Oryza sativa
; STRAIN: Akihikari
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..2845
; OTHER INFORMATION: /label= PT72
; OTHER INFORMATION: /note= "sequence comprising anther specific
; OTHER INFORMATION: promoter PT72"

FEATURE:
NAME/KEY: - 2733..2739
LOCATION: 2733..2739 /label= TATA
OTHER INFORMATION: /note= "TATA Box"
FEATURE:
NAME/KEY: -
LOCATION: 2765 /note= "transcription initiation
OTHER INFORMATION: determined by primer extension"
FEATURE:
NAME/KEY: - 2846..2848
LOCATION: 2846..2848 /label= ATG
OTHER INFORMATION: /note= "ATG start of translation of rice T72 gene"
US-08-351-413-7

Query Match 9.2%; Score 35.6; DB 1; Length 3627;
Best Local Similarity 47.0%; Pred. No. 0.39;
Matches 110; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 1 GACTTCAAAGCAAGCTGTATTTTCAACAAATCTTCTAATTCCTGTGTGCCAGGCA 60
DB 1188 GAGAGCAATGCAATGTCACAAACAAACAGGCTCTTCCAAACGGTGTGTTTCATTCCA 1247

QY 61 GGGAGACGGTTTCCAGGAGGGGCGGCCCTGTGTGCGAGTTCGGATGTTATTAGATGTT 120
DB 1248 CAGAACAGGATAGCAACAGAGCAACAAACCGTTCAACAATATATATATATATATAT 1307

QY 121 ACAAGTTAT 180
DB 1308 AT 1367

QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCCTGGACTGCTAC 234
DB 1368 ATACGGAACCTTAAACACATGAAATCTAAACATTTTCAACCAATCAGAACTAC 1421

RESULT 14

US-09-025-583-7
Sequence 7, Application US/09025583
Patent No. 5977433
GENERAL INFORMATION:
APPLICANT: Williams, Mark
APPLICANT: Leemans, Jan
TITLE OF INVENTION: Maintenance of male-sterile plants
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESS: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 2046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/025,583
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/351,413
FILING DATE: US 07/899,072
FILING DATE: 12-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,849
FILING DATE: 03-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Svensson, Leonard R.

REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 2121-102PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3627 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Oryza sativa
STRAIN: Akihikari
FEATURE:
NAME/KEY: - 1..2845
LOCATION: 1..2845 /label= PT72
OTHER INFORMATION: /note= "sequence comprising anther specific
OTHER INFORMATION: promoter PT72"
FEATURE:
NAME/KEY: -
LOCATION: 2733..2739 /label= TATA
OTHER INFORMATION: /note= "TATA Box"
FEATURE:
NAME/KEY: -
LOCATION: 2765 /note= "transcription initiation
OTHER INFORMATION: determined by primer extension"
FEATURE:
NAME/KEY: - 2846..2848
LOCATION: 2846..2848 /label= ATG
OTHER INFORMATION: /note= "ATG start of translation of rice T72 gene"
US-09-025-583-7

Query Match 9.2%; Score 35.6; DB 2; Length 3627;
Best Local Similarity 47.0%; Pred. No. 0.39;
Matches 110; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 1 GACTTCAAAGCAAGCTGTATTTTCAACAAATCTTCTAATTCCTGTGTGCCAGGCA 60
DB 1188 GAGAGCAATGCAATGTCACAAACAAACAGGCTCTTCCAAACGGTGTGTTTCATTCCA 1247

QY 61 GGGAGACGGTTTCCAGGAGGGGCGGCCCTGTGTGCGAGTTCGGATGTTATTAGATGTT 120
DB 1248 CAGAACAGGATAGCAACAGAGCAACAAACCGTTCAACAATATATATATATATATAT 1307

QY 121 ACAAGTTAT 180
DB 1308 AT 1367

QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCCTGGACTGCTAC 234
DB 1368 ATACGGAACCTTAAACACATGAAATCTAAACATTTTCAACCAATCAGAACTAC 1421

RESULT 15

US-10-204-708-67
Sequence 67, Application US/10204708
Patent No. 667731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
TITLE OF INVENTION: by Assessing DNA Methylation
FILE REFERENCE: 5013.1012


```
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 67
; LENGTH: 6368
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-67

Query Match          9.1%; Score 35.2; DB 4; Length 6368;
Best Local Similarity 49.0%; Pred. No. 0.7;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 114 AGATGTTACAGTTTATATATCTATATATATAATTTATTTAGTGTTCACAGATGTAT 173
Db 114 AGTTTATTTAAGTTTATATATATGTTTGTGATAGATATATATTTTGTAGGGGGTT 4173

QY 174 TTGTTGTAGACTTAACACTTCTTACGCAATGCTCTAGAGTTTATAGCCTGGACTGCTA 233
Db 174 AGTTTATTTTAAATATATAGATTATGTAGTGTGTGAAATTTTAAAGTTATATAGTA 4233

QY 234 CCTTCAAAGCTTGGAGGGAGCCGCAATTCAGTTGGTTCTGTCTGTACTGTACTGGG 293
Db 4234 GGTTTATAGGGTTTAAAGTAGAGGAGTCAGTTTATTTCCGGAAGGTTTAGAGTAGTTT 4293

QY 294 CCTGAGTCTGG 305
Db 4294 TAGTAATTACGG 4305
```

Search completed: September 22, 2004, 12:33:21
Job time : 78 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 22, 2004, 10:37:05 ; Search time 87.5 Seconds
(without alignments)
2486.418 Million cell updates/sec

Title: M64347

Perfect score: 702

Sequence: 1 GACTTCAAAGCAAGCTGCTA.....GTGCATGTGGCCAGAGGTG 385

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q/cgn2 1/USPTO.spool p/CANELLA305B/runat 22092004 112334 14487/app query.fasta_1.583
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SURFIX=rag -MINMATCH=0_1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=CANELLA305B @CGN 1 1 224 @runat 22092004 112334 14487 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	94	13.4	561	AAU02978	Angiotensin
2	83	12.0	494	ABG12606	Novel hum
3	83	12.0	504	AA95722	Human pro
4	76	11.0	381	2 AA72063	HSV-2 str
5	75	10.9	966	5 AAU80189	Human TSP
6	72.5	10.3	147	5 ABB90106	Human pol
7	71	10.3	961	5 AAU80188	Human TSP
8	71	10.1	1045	6 ABO07129	Novel hum
9	71	10.1	2273	7 ADCS1662	Human MEG
10	71	10.1	2778	7 ADCS1660	Human MEG

11	70.5	10.0	1043	7	ADD04782	Adg04782 Pig RAG-1
12	70.5	10.2	2130	3	RAG50617	Arabidops
13	70.5	10.2	2138	3	RAG50616	Arabidops
14	70.5	10.2	2204	3	RAG50615	Arabidops
15	70.5	10.2	2254	2	AA776949	Accase. 2
16	70.5	10.2	2254	2	AA722129	Arabidops
17	70.5	10.2	2254	2	AA740598	Arabidops
18	70	10.1	865	4	AB644483	Drosophil
19	70	10.1	870	3	AB21252	Rat metal
20	69	10.0	420	6	ABR42500	Coumeryc
21	68.5	9.9	93	5	ABP64921	Human pro
22	68.5	9.8	363	2	AA737863	Amino aci
23	68.5	9.8	391	7	ADC87517	Human GPC
24	68.5	9.8	1040	2	AAW73584	RAG-1 pro
25	68.5	9.9	1186	4	AB63516	Drosophil
26	68	9.9	168	6	ADA83946	Human POM
27	68	9.9	420	6	ABR42541	Clorobloc
28	67.5	9.6	141	3	ABBS3247	Human col
29	67.5	9.8	396	7	ADB37611	Neural th
30	67.5	9.8	431	2	AA750027	Human dim
31	67.5	9.8	475	2	AA734000	Human dim
32	67.5	9.6	526	5	ABG30696	Murine md
33	67.5	9.6	527	5	AB80743	Rat YAK1
34	67.5	9.8	599	6	AAE37108	Human sel
35	67.5	9.8	599	6	ABR42589	Selective
36	67	9.5	326	5	ABBS7350	Mouse isc
37	67	9.5	528	2	AAW64559	Human pro
38	67	9.5	528	4	AAE02011	Human YAK
39	67	9.5	528	7	ABR57642	Human DYR
40	67	9.5	549	3	AA768781	Amino aci
41	67	9.7	628	7	ADB85485	Human agg
42	67	9.7	752	7	ADB85483	Human agg
43	67	9.7	753	7	ADB85481	Human agg
44	67	9.7	763	7	ADB85492	Human agg
45	67	9.7	930	2	AAW75426	Human agg

ALIGNMENTS

RESULT 1

AAU02978

ID AAU02978 standard; protein; 561 AA.

XX AAU02978;

AC AAU02978;

XX 12-SEP-2001 (first entry)

XX Angiotensin converting enzyme (ACEV) splice variant protein #78.

XX Angiotensin converting enzyme splice variant; ACEV; interleukin 6;
XX granulocyte colony stimulating factor receptor; glucagon; hypertrophy;
XX platelet-derived endothelial cell growth factor; cardiovascular disease;
XX cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1C;
XX vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer;
XX myocardial infarction; coronary arterial thrombosis; renal disease;
XX diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;
XX multiple sclerosis; immune complex nephritis; deep vein thrombosis;
XX noncardiotoxic pulmonary granulomatous disease; endothelial abnormality;
XX vascular disorder; asbestosis.

OS Homo sapiens.

XX WO200136632-A2.

XX 25-MAY-2001.

XX 17-NOV-2000; 2000MO-IL000766.

XX 17-NOV-1999; 99IL-00132978.

XX 10-DEC-1999; 99IL-00133455.

XX (COMP-) COMPUGEN LTD.

Db 279 SerAsnTyrSerThrProArg 285
 RESULT 5
 AAU80189
 ID AAU80189 standard; protein; 966 AA.
 XX AC
 XX AAU80189;
 XX DT
 XX 15-JUL-2002 (first entry)
 XX DE Human TSPI domain containing protein encoded by cDNA FG01869.
 XX KW TSPI; thrombospondin domain; FG01869; angiogenesis; vasculogenesis;
 XX KW human.
 XX OS Homo sapiens.
 XX FN JP2002085059-A.
 XX PD 26-MAR-2002.
 XX PF 08-SEP-2000; 2000JP-00273778.
 XX XX 08-SEP-2000; 2000JP-00273778.
 XX (KAZU-) ZH KAZUSA DNA KENKYUSHO.
 XX PA (YOSH) YOSHITOMI PHARM IND KK.
 XX DR WPI; 2002-378268/41.
 XX DR N-PSDB; ABK50390.
 XX TSPI domain-containing polypeptide useful for drug compositions.
 PT Claim 1; Page 30-33; 51pp; Japanese.
 XX The invention relates to a TSPI (thrombospondin 1) domain-containing
 CC polypeptide comprising the proteins appearing as AAU80188 and AAU80189,
 CC encoded by cDNAs designated FG03969 and FG01896. Also included are
 CC proteins that are 50% homologous to the proteins and a polypeptide having
 CC at least one deletion, replacement, addition or insertion of amino acid
 CC in the proteins and having at least 8 repetitions of the TSPI domain. The
 CC polypeptide can be used in drug compositions particularly for disorders
 CC associated with angiogenesis and vasculogenesis. The present sequence is
 CC the TSPI domain containing protein encoded by cDNA FG01869
 XX
 SQ Sequence 966 AA;
 Alignment Scores:
 Pred. No.: 3.3 Length: 966
 Score: 75.00 Matches: 41
 Percent Similarity: 29.79% Conservative: 15
 Best Local Similarity: 21.81% Mismatches: 51
 Query Match: 10.87% Indels: 81
 DB: 5 Gaps: 9
 M64347 (1-385) x AAU80189 (1-966)
 QY 385 CACTCTGGCCACATGCTGGCCGCCAAGAGAGACACACCTGAGCCATGCCCTGCA 326
 Db 460 HisArgSerProCysLeuGlyProAspThrGlnThrArgGlnGlnProCysProGly 479
 QY 325 -----GGCAAGCAAGGACAGCTGCCAGACTCCAGCTCAGGCCCCA--- 290
 Db 480 LeuLeuGluAlaCysSerTrpGlyProTyrProCysSerArgSerGlyProGly 499
 QY 289 -----GTACAGTACAGACCAACCAACTGATTC 260
 Db 500 LeuAlaSerArgSerGlySerCysProCysLeuMetAlaAlaAspProThrCysAsn 519
 QY 259 ACGGCTTCCCTCCAA-----GCTTTGAAAGGTAGCTCAGGCTATAAA 215
 Db 520 SerThrPheLeuHisLeuAspThrGlnGlyCystyrSerGlyProCysPro----- 536

QY 214 ACTCTAGAAGCATTGCGTA---AGAAGTGTTAAGTCTACACAAATACATCTTGTAAAAA 158
 Db 537 -----Glu-GluCysValTrpSerSerTrpSerSerTrpArgCysSerCysArgVa 554
 QY 157 CTCAATAAATTAT 98
 Db 554 lLeuValGlnGlnArgTyrArg----- 561
 QY 97 GCACACAGGCGCGCC----- 82
 Db 562 -HisGlnGlyProAlaSerArgGlyAlaArgAlaGlyAlaProCysThrArgLeuAspG1 581
 QY 81 -----CCTCCTCGAA 71
 Db 581 yHisPheArgProCysLeuLeuSerAsnCysSerGluAspSerCysThrProProPheG1 601
 QY 70 A-----CGGTCTCCCTGCTGGGACACACAGCAATT-----AGAAGAAT 32
 Db 601 upheHisAlaCysGlySerProCysAlaGlyLeuCysAlaThrHisLeuSerHisGlnLe 621
 QY 31 TTGTATGAAATACACAGCTTGC 10
 Db 621 uCysGlnAspLeuProProCys 628
 RESULT 6
 ABB90106
 ID ABB90106 standard; protein; 147 AA.
 XX AC ABB90106;
 XX XX
 XX DT 24-MAY-2002 (first entry)
 XX DE Human polypeptide SEQ ID NO 2482.
 XX Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
 KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiact; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 OS Homo sapiens.
 XX WO200190304-A2.
 XX PD 29-NOV-2001.
 XX PF 18-MAY-2001; 2001WO-US016450.
 XX PR 19-MAY-2000; 2000US-0205515P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX PI Birse CE, Rosen CA;
 XX WPI; 2002-122018/16.
 XX N-PSDB; ABL90515.
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders.
 PS Claim 11; SEQ ID NO 2482; 2081pp + Sequence Listing; English.
 XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABB9040-ABB90444) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (anti)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic

anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
cardiovascular disorders such as myocardial ischaemias; (d) wound healing
(e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
infectious diseases such as viral, bacterial, fungal and parasitic
infections. Note: The sequence data for this patent did not form part of
the printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published/pct/sequences

Sequence 147 AA;

Alignment Scores:	4.13	Length:	147
Pred. No.:	72.50	Matches:	34
Score:	42.00%	Conservative:	19
Percent Similarity:	26.98%	Mismatches:	40
Best Local Similarity:	10.33%	Indels:	33
Query Match:	5	Gaps:	8
DB:			

M64347 (1-385) x ABB90106 (1-147)

QY	43	TGCTGTGTCCTCCAGCGAGGAGACGGTTTCCAGGG-----AGGGCCGGGCC-----	90	
Db	9	:::		
QY	91	-----TGTCGACGGT-----CCGATGTTATAGACTTACAAGTTTATATATA 135		
Db	29	:::		
QY	136	TCTATATATAAATTATTGAGTTTTACAAGATGATTGTTGTAGACTTAACACTTCT 195		
Db	49	:::		
QY	196	TACGCAATGCTTCTAGACTTTATAGCCTGGACTGCTACCTTTCAAAGCTTGGAGGAAAG 255		
Db	60	:::		
QY	255	CGGTGNAATCAGTTGGTTCGTCTGTACTGTTACTGGGCCCTGAGTCT-----GGGCGAG 309		
Db	74	:::		
QY	310	CTGTGCTCTGCTGCTGCAGGCGC-----CATGGCTCAGAGGTGG 348		
Db	94	:::		
QY	349	TCTCTTCTTTGGGCCCGAG 366		
Db	114	SergIvLeuGlyLeuGln 119		

RESULT 7
AAU80188
ID AAU80188 standard; protein; 961 AA.

XX
AC
AAU80188:

XX 15-JUL-2002 (first entry)

Human TSP1 domain containing XX

TS1; thrombospondin domain; FG06969; angiogenesis; vasculogenesis;
human.

XX
OS
Homo sapiens.XX
PN JP2002085059-A.

XX PD 26-MAR-2000

XX
PF 08-SEP-2000: 2000TP-0027377808-SEP-2000 2000TP-00273778
XX
PP

XX
PA (KAZU-) ZH KAZUSA DNA KENKYUSHO

PA (YOSH) YOSHITOMI PHARM IND KK.

XX
DR WPI: 2002-378268/41.XX
DR N-PSDB; ABX50389.

PT TSPI domain-containing polypeptide useful for drug compositions.

PS
vv
Claim 1; Page 23-26; 51pp; Japanese.

CC The invention relates to a TSP1 (thrombospondin 1) domain-containing
CC polypeptide comprising the proteins appearing as AAU80188 and AAU80189,
CC encoded by cDNAs designated FGO6969 and FGO1896. Also included are
CC proteins that are 50% homologous to the proteins and a polypeptide having
CC at least one deletion, replacement, addition or insertion of amino acid
CC in the proteins and having at least 8 repetitions of the TSP1 domain. The
CC polypeptide can be used in drug compositions particularly for disorders
CC associated with angiogenesis and vasculogenesis. The present sequence is
CC the TSP1 domain containing protein encoded by cDNA FGO6969

Sequence 961 AA:

Alignment Scores:	
Pred. No.:	10.9
Score:	71.00
Percent Similarity:	33.33%
Best Local Similarity:	23.33%
Query Match:	10.29%
DB:	5
Length:	961
Matches:	35
Conservative:	15
Mismatches:	39
Indels:	61
Gaps:	9

M64347 (1-385) x AAT80188 (1-961)

385	QY	CACCTCTGGCCACCATGCACCTGGGCCCCAAGAAGACAGACACCCCTGAGCCATGGCCCTCGCA	326
511	Db	HisArgSerProProCysLeuGlyProAspThrGlnThrArgGlnGlnProCysProGly	530
325	QY	-----GGCAACGAGGACAGCTGCCAGACTCAGGGCCACGTA	287
531	Db	LeuLeuGluAlaCysSerTrpGlyProTrpGlyProCysSerArgSerCysGlyPro	549
286	QY	ACAGTAGACAAGCAACCAACTGAATTCACGGCTTCCTCCAAGCTTTGAAAGGTAGACAGT	227
550	Db	-----GlyLeuAlaSerArgSerGlySerCys	558
226	QY	CCAGCTATAAACTCTAGAGCAATTCGGTAGAAG-----TGTTAAGTCTAC	179
559	Db	Pro-----CysLeuMetAlaLysAlaAspProThrCys	569
178	QY	AACAAATACATCTTGTAATAAACTCAATAAAATTATATATAGATATATATAAACTTGTA	119
570	Db	-----AsnSerThrPheLeu-----	574
118	QY	CATCTAATAACATCGAACCTGCACAGGGCG-----GCCCTCC	77
575	Db	HisLeuAspThr-GlnGlyCysTrpSerGlyProCysProGluAspSerCysThrProPr	594
76	QY	CTGGAAA-----CGCTCTCCCTCGGACACACAGCAATT-----AG	38
594	Db	ophegldpHeHisAlaCysGlySerProCysAlaGlyIleCysAlaThrHisLeuSerHi	614
37	QY	AAGAATTGTATGAAATACACAGCTTGC	10
614	Db	sglnLeuCysGlnAspLeuProCys	623

RESULT 8

ABO07129
ID ABO

XX
AC
AB007129.

XX
DT 12-2170-30

XX
DE
Nov-07 human protocol NOV 11 07

XX

CC NOVX; autoimmune disease; allergy; Alzheimer's disease; stroke;
 CC Parkinson's disease; Huntington's disease; multiple sclerosis; addiction;
 CC anxiety; pain; diabetes; glomerulonephritis; obesity;
 CC systemic lupus erythematosus; asthma; scleroderma; pancreatitis;
 CC graft versus host disease; ulcer; anaemia; cancer; trauma; infection;
 CC cardiomyopathy; atherosclerosis; hypertension; AIDS; Crohn's disease;
 CC acquired immunodeficiency syndrome; chromosomal mapping; tissue typing;
 CC forensic biology; predictive medicine; gene therapy; human.

XX Homo sapiens.

OS WO200298900-A2.

PN 12-DEC-2002.

PD 04-JUN-2002; 2002WO-US017558.

PF 04-JUN-2001; 2001US-0295607P.

PR 04-JUN-2001; 2001US-0295661P.

PR 06-JUN-2001; 2001US-0296404P.

PR 06-JUN-2001; 2001US-0296418P.

PR 07-JUN-2001; 2001US-0296575P.

PR 11-JUN-2001; 2001US-0297414P.

PR 12-JUN-2001; 2001US-0297567P.

PR 15-JUN-2001; 2001US-0298528P.

PR 18-JUN-2001; 2001US-0299133P.

PR 19-JUN-2001; 2001US-0299230P.

PR 21-JUN-2001; 2001US-0300177P.

PR 22-JUN-2001; 2001US-0300893P.

PR 26-JUN-2001; 2001US-0301530P.

PR 28-JUN-2001; 2001US-0301550P.

PR 03-JUL-2001; 2001US-0302951P.

PR 12-SEP-2001; 2001US-0318727P.

PR 27-SEP-2001; 2001US-0325685P.

PR 22-FEB-2002; 2002US-0358814P.

PR 03-JUN-2002; 2002US-00161927.

XX (CURA-) CURAGEN CORP.

XX Zerhusen BD, Kekuda R, Spytek KA, Shenoy SG, Miller CE, Hjalt T;

PI Gerlach VL, Baumgartner JC, Guo X, Gangolli EA, Vernet CAM;

PI Padigar M, Li L, Pena CEA, Gorman L, Anderson DW, Edinger SR;

PI Patturajan M, Stone DJ;

XX WPI; 2003-140585/13.

DR N-PSDB; ACD13203.

XX Novel isolated NOVX polypeptide useful treating or preventing disorders
 or syndromes such as autoimmune disease, allergies, Alzheimer's disease,
 stroke, Parkinson's disease, Huntington's disease or multiple sclerosis.

PS Claim 1; Page 144; 408pp; English.

CC medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical
 CC trials and pharmacogenomic), and in methods of treatment (e.g.,
 CC therapeutic and prophylactic). (ii) is useful in gene therapy, to express
 CC (i), to detect NOVX mRNA or a genetic lesion in a NOVX gene, and to
 CC modulate NOVX activity. This is the amino acid sequence of a novel human
 CC NOV protein

XX Sequence 1045 AA;

Alignment Scores:

Pred. No.:	11.1	Length:	1045
Score:	71.50	Matches:	24
Percent Similarity:	40.24%	Conservative:	9
Best Local Similarity:	29.27%	Mismatches:	25
Query Match:	10.11%	Indels:	24
DB:	6	Gaps:	5

M64347 (1-385) x ABO07129 (1-1045)

QY 189 CACTTCTTACCAATGCTTCTAGAGTTTATAGCTGCTGCTGCTACCT-----TTCAAA 242

Db 225 HisArgLeuHisAsnSerValSerAlaProSerProGlyIleProProArgAspPheLys 244

QY 243 GCT-----TGGAGGGAAGCGGTGAATCA 266

Db 245 SerLeuAlaLeuAlaArgAlaProGlyHisGlyGlyPheTrp---GinGlyValAlaAla 263

QY 267 GTTGGTTCTGTTCTCTACTGTTACTGGG-----CCCTGAGTCTGGGAGCTGTC 314

Db 264 GluGlyValGlyCysThrLeuThrGlyAlaTrpArgSerProValProTrpSerGlyThr 283

QY 315 CTTGCTTGGCTGCA-----GGGCCATGGCTCAGGGTGGTCTCTCTCTGG 359

Db 284 GlyCysValProGlyGlyPheThrValProGlyProArgProAlaProAlaProTrp 303

QY 360 GGCCCA 365

Db 304 GlyPro 305

RESULT 9

ADCS1662

ID ADCS1662 standard; protein; 2273 AA.

AC ADCS1662;

DT 18-DEC-2003 (first entry)

DE Human MEGF8 protein #2.

XX Human; membrane binding-MEGF8; secretory-MEGF8; plexin domain;

XX axial fibre induction; nerve cell; heart development;

XX skeleton development; immune response; vasculogenesis; cancer growth;

XX metastasis; cytostatic; neuroprotective; cardiant; immunomodulatory;

XX vulnerable.

XX Homo sapiens.

OS JP2002360254-A.

PN 17-DEC-2002.

PD 21-MAY-2001; 2001JP-00151059.

PF 27-MAR-2001; 2001JP-00090438.

PR (KAZU-) ZH KAZUSA DNA KENKYUSHO.

XX WPI; 2003-516152/49.

XX N-PSDB; ADCS1661.

XX A membrane binding- or secretory-MEGF8 gene and a protein encoded by the

PT gene, for diagnosis and treatment of diseases caused by plexin domain.

XX


```

PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145445P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145313P.
PR 27-JUL-1999; 99US-0145318P.
PR 27-JUL-1999; 99US-0145319P.
PR 28-JUL-1999; 99US-0145351P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 31-AUG-1999; 99US-0151303P.
PR 01-SEP-1999; 99US-0151438P.
PR 07-SEP-1999; 99US-0151930P.
PR 10-SEP-1999; 99US-0152363P.
PR 13-SEP-1999; 99US-0153070P.
PR 15-SEP-1999; 99US-0153758P.
PR 16-SEP-1999; 99US-0154018P.
PR 20-SEP-1999; 99US-0154039P.
PR 22-SEP-1999; 99US-0154779P.
PR 23-SEP-1999; 99US-0155132P.
PR 24-SEP-1999; 99US-0155486P.
PR 28-SEP-1999; 99US-0155659P.
PR 29-SEP-1999; 99US-0156458P.
PR 04-OCT-1999; 99US-0156596P.
PR 05-OCT-1999; 99US-0157117P.
PR 06-OCT-1999; 99US-0157753P.
PR 07-OCT-1999; 99US-0157865P.
PR 08-OCT-1999; 99US-0158029P.
PR 12-OCT-1999; 99US-0158332P.
PR 13-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.

PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 26-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 28-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 15.7 Length: 2130
Score: 70.50 Matches: 30
Percent Similarity: 38.53% Conservative: 12
Best Local Similarity: 27.52% Mismatches: 40
Query Match: 10.22% Indels: 27
DB: 3 Gaps: 4

M64347 (1-385) x AAG50616 (1-2130)
QY 364 GGGCCCAAGAGAGACCCCTGAGC-----CATGCCCTGCAGGC-----AAG 320
DB 1393 GlyProLeuHisGluThrProIleSerAspGlnTyrLysProLeuGlyTyrLeuAspArg 1412
QY 319 CAAGGCACAGCTGCCAGACTCAGGCGCCAGTACAGTACAGACGACCACTCAATTC 260
DB 1413 GlnArgLeuAlaAlaArgSerAsnThrThrTyrCysTyrAspPheProLeuAlaPhe 1432
QY 259 ACGGCTTCCTCCCAAGCTTTCAGAGTAGCAGT---CCAGGCTATAAACTCTAGAAGCA 203
DB 1433 GlyThrAlaLeuGluLeuLeuTyrPAlaSerGlnHisProGlyValLys----- 1448
QY 202 TTGGTAAGAAGTGTAAAGTCTACACAAATACATCTGTAAAAACTCAATAAATTATAT 143
DB 1449 -----LysProTyr 1451
QY 142 ATATAGATATATAAATCTGTAATCTATAATACATCGGAACCTGCACACAGGCGCGGC 83
DB 1452 LysAspThrLeuIleAsnValLysGluLeuValPheSerLysProGluGlySerSerdgly 1471
QY 82 CCCTCCCTGGAAACCGTCTCCCTGCCT 56
DB 1472 ThrSerLeuAspLeuValGluArgPro 1480

RESULT 13
AAG50616
ID AAG50616 standard; protein; 2158 AA.
XX
AC AAG50616;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 64164.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX

```

[illegible]

PR	20-SEP-1999;	99US-0154779P.	Db	1500 ThrSerLeuAspLeuValGluArgPro 1508
PR	22-SEP-1999;	99US-0155139P.	RESULT 14	
PR	23-SEP-1999;	99US-0155486P.	AAG50615	
PR	24-SEP-1999;	99US-0155659P.	ID	AAG50615 standard; protein; 2204 AA.
PR	28-SEP-1999;	99US-0156458P.	XX	AAG50615;
PR	29-SEP-1999;	99US-0156596P.	AC	
PR	04-OCT-1999;	99US-0157117P.	XX	18-OCT-2000 (first entry)
PR	06-OCT-1999;	99US-0157553P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	07-OCT-1999;	99US-0157865P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	08-OCT-1999;	99US-0158029P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	08-OCT-1999;	99US-0158232P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	12-OCT-1999;	99US-0158365P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	13-OCT-1999;	99US-0159293P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	13-OCT-1999;	99US-0159294P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	13-OCT-1999;	99US-0159295P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	14-OCT-1999;	99US-0159330P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	14-OCT-1999;	99US-0159331P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	14-OCT-1999;	99US-0159637P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	14-OCT-1999;	99US-0159638P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	18-OCT-1999;	99US-0159584P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	21-OCT-1999;	99US-0160741P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	21-OCT-1999;	99US-0160767P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	21-OCT-1999;	99US-0160768P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	21-OCT-1999;	99US-0160770P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	21-OCT-1999;	99US-0160814P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	21-OCT-1999;	99US-0160815P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	22-OCT-1999;	99US-0160980P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	22-OCT-1999;	99US-0160981P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	22-OCT-1999;	99US-0160989P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	25-OCT-1999;	99US-0161404P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	25-OCT-1999;	99US-0161405P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	25-OCT-1999;	99US-0161406P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	26-OCT-1999;	99US-0161359P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	26-OCT-1999;	99US-0161360P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	26-OCT-1999;	99US-0161361P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	28-OCT-1999;	99US-0161920P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	28-OCT-1999;	99US-0161923P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	28-OCT-1999;	99US-0161933P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
PR	29-OCT-1999;	99US-0162142P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 64163.
Alignment Scores:			25-FEB-2000; 2000EP-00301439.	
Pred. No.:	15.8	Length:	2158	
Score:	70.50	Matches:	30	
Percent Similarity:	38.53%	Conservative:	12	
Best Local Similarity:	27.52%	Mismatches:	40	
Query Match:	10.22%	Indels:	27	
DB:	3	Gaps:	4	
M64347 (1-385) x AAG50616 (1-2158)			25-FEB-1999; 99US-0121825P.	
QY	364	GGGCCCCAAGAGAGACCCCTGAGC-----CATGGCCCTGCAGGC-----AAG	320	05-MAR-1999; 99US-0123180P.
Db	1421	GlyProLeuHisGluThrProLeuSerAspGlnTyrLeuLeuAspArg	1440	09-MAR-1999; 99US-0123548P.
QY	319	CAAGGACAGCTCCAGACTCAGGCCAGTACAGACGACCACTGAATTC	260	23-MAR-1999; 99US-0125788P.
Db	1441	GlnArgLeuAlaAlaArgSerAsnThrThrTyrCysTyrAspPheProLeuAlaPhe	1460	25-MAR-1999; 99US-0126684P.
QY	259	ACGGTTCCTCCAGCTTGAAGTAGCAGT---CCAGGCTATAAACTCTAGAGCA	203	01-APR-1999; 99US-0126785P.
Db	1461	GlyThrAlaLeuGluLeuTyrAlaSerGlnHisProGlyVallys	1476	08-APR-1999; 99US-0128234P.
QY	202	TTGCGTGAAGAAGTGTAAAGCTACAAATACATCTGTAAAACTCAATAATATATAT	143	16-APR-1999; 99US-0128714P.
Db	1477	-----LysProTyr	1479	19-APR-1999; 99US-0130077P.
QY	142	ATATAGATATATATACTTGAACATCTAATACATCGAACCTCACACAGGCGCGC	83	21-APR-1999; 99US-0130449P.
Db	1480	LysAspThrLeuIleAsnValysGluLeuValPheSerTyrProGluGlySerGly	1499	23-APR-1999; 99US-0130891P.
QY	82	CCCTCCCTGGAACCGTCTCCCTGCCT	56	28-APR-1999; 99US-0131449P.

```

PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142390P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144894P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.

PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156598P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158039P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159233P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 15.9
Score: 70.50
Percent Similarity: 38.53%
Best Local Similarity: 27.52%
Query Match: 10.22%
DB: 3

M64347 (1-385) x AAG50615 (1-2204)

Qy 364 GGGCCCCAGAGAGACCCCTGAGC-----CATGGCCCTGAGC-----AAG 320
Db 1467 GlyProLeuHisGlnThrProIleSerAspGlnTy-LysProLeuGlyTyLeuAspArg 1486

```

```
QY 319 CAGGGACAGCTGCCAGACTCAGGGCCAGTACAGTACAGAACGACCAACTGAATTC 260
Db |||||
QY 1487 GlnArgLeuAlaAlaArgSerAsnThrThrTyrCysTyrAspPheProLeuAlaPhe 1506
Db |||||
QY 259 ACGGCTTCCTCCCAAGCTTTGAAAGTAGCAGT---CCAGGCTATAAAACTCTAGAAGCA 203
Db :|||:
QY 1507 GlyThrAlaLeuGluLeuTrpAlaSerGlnHisProGlyVallys----- 1522
QY 202 TTGCGTAAGAGTGTAAAGTCTACAAATAACATCTTGTAAAACTCAATAAATTATAT 143
Db |||||
QY 1523 -----LysProTyr 1525
QY 142 ATATAGATATATATAAAGTTGTAACATCTATAATACATCGGAACCTGCACACAGGCGCGGC 83
Db |||||
QY 1526 LysAspThrLeuIleAsnVallysGluLeuValPheSerLysProGluGlySerSerGly 1545
Db :|||:
QY 82 CCCTCCCTGGAAACCGCTCCCTGCCT 56
Db |||||
QY 1546 ThrSerLeuAspLeuValGluArgPro 1554
Db |||||

RESULT 15
AAR76949
ID AAR76949 standard; protein; 2254 AA.
XX
AC AAR76949;
DT
DT 28-FEB-1996 (first entry)
XX
DE ACCase.
XX
KW Polymerase chain reaction; PCR; primer; amplify; acetyl CoA carboxylase;
KW ACCase; transgenic plant; regulation; fat; protein; agricultural plant.
XX
OS Arabidopsis thaliana.
XX
XX
XX JP07143887-A.
XX
XX 06-JUN-1995.
XX
XX 28-JUN-1994; 94JP-00146827.
XX
XX 17-AUG-1993; 93JP-00203477.
XX
XX (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.
XX
XX WPI; 1995-236465/31.
XX
XX N-PSDB; AAQ93232.
XX
XX Plant acetyl CoA carboxylase gene obtd. by PCR amplification - useful for
PT increasing fat/protein content in plants.
XX
XX Claim 11; Page 14-22; 33pp; Japanese.
XX
XX This sequence is encoded by the cDNA of the acetyl CoA carboxylase
CC (ACCcase) gene which was amplified using the primers given in AAQ93221- 30
CC and AAQ93233-36. This full length ACCase gene may be used to produce a
CC transgenic plant such that it expresses a changed amount of ACCase. The
CC ACCase gene may be used to regulate the production of fat/protein in
CC agricultural plants
XX
XX SQ Sequence 2254 AA;

Alignment Scores:
Pred. NO.: 16 Length: 2254
Score: 70.50 Matches: 30
Percent Similarity: 38.53% Conservative: 12
Best Local Similarity: 27.52% Mismatches: 40
Query Match: 10.22% Indels: 27
DB: 2 Gaps: 4

M64347 (1-385) x AAR76949 (1-2254)
QY 364 GGGCCCCAAGAGAGACCACTGAGC-----CATGGCCCTGCAGGC-----AAG 320
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 22, 2004, 10:47:46 ; Search time 23.5 Seconds
(without alignments)
1691.574 Million cell updates/sec

Title: M64347

Perfect score: 702

Sequence: 1 GACTTCAAGCAGCTGTGTA.....GTGCATGTGGCCAGAGGTG 385

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/CNELLIA305B/runat_22092004.112337.14559/app_query.fasta_1.583
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -DOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=CNELLIA305B @CGN 1.1 46 @runat 22092004.112337.14559 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/58 COMB pep:*
- 2: /cgn2_6/prodata/2/iaa/58 COMB pep:*
- 3: /cgn2_6/prodata/2/iaa/6A COMB pep:*
- 4: /cgn2_6/prodata/2/iaa/6B COMB pep:*
- 5: /cgn2_6/prodata/2/iaa/PTCUS COMB pep:*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	10.7	539	US-09-252-991A-17070	Sequence 17070, A
C 2	70.5	10.2	2254	US-08-677-010-3	Sequence 3, Appli
C 3	70.5	10.2	2254	US-08-790-519-3	Sequence 3, Appli
4	70	10.0	205	US-09-252-991A-16946	Sequence 16946, A
5	68.5	9.8	1040	US-08-254-989-2	Sequence 2, Appli
6	67.5	9.6	527	US-09-659-166-2	Sequence 2, Appli
7	67	9.5	528	US-08-802-466-2	Sequence 2, Appli
8	67	9.5	528	US-09-350-484-2	Sequence 2, Appli
C 9	67	9.7	930	US-09-122-126B-15	Sequence 15, Appl
C 10	67	9.7	930	US-09-634-286A-15	Sequence 15, Appl
C 11	66.5	9.6	2523	US-08-185-432-18	Sequence 18, Appl
C 12	66.5	9.6	2523	US-08-899-233-3	Sequence 3, Appli

C 13	65.5	9.5	140	4	US-09-252-991A-30876	Sequence 30876, A
C 14	65	9.4	572	4	US-09-252-991A-21105	Sequence 21105, A
C 15	65	9.4	930	4	US-09-369-364A-2	Sequence 2, Appli
C 16	64.5	9.3	160	4	US-09-252-991A-31945	Sequence 31945, A
C 17	64	9.3	154	4	US-08-489-039A-9332	Sequence 9332, Ap
C 18	64	9.3	166	4	US-09-252-991A-21902	Sequence 21902, A
C 19	64	9.3	305	4	US-10-162-012-23	Sequence 23, Appl
C 20	64	9.3	304	4	US-09-976-594-615	Sequence 615, App
21	64	9.1	1214	2	US-08-231-193A-54	Sequence 54, Appl
22	64	9.1	1214	2	US-08-486-273A-54	Sequence 54, Appl
23	64	9.1	1214	3	US-08-480-474-54	Sequence 54, Appl
24	64	9.1	1214	3	US-08-940-086A-54	Sequence 54, Appl
25	64	9.1	1214	4	US-08-940-035A-54	Sequence 54, Appl
26	64	9.1	1214	4	US-08-935-105A-54	Sequence 54, Appl
27	64	9.1	1214	4	US-09-648-797-54	Sequence 54, Appl
28	64	9.1	1214	4	US-09-386-123-54	Sequence 54, Appl
29	64	9.1	1219	2	US-08-231-193A-50	Sequence 50, Appl
30	64	9.1	1219	2	US-08-486-273A-50	Sequence 50, Appl
31	64	9.1	1219	3	US-08-480-474-50	Sequence 50, Appl
32	64	9.1	1219	3	US-08-940-086A-50	Sequence 50, Appl
33	64	9.1	1219	4	US-08-940-035A-50	Sequence 50, Appl
34	64	9.1	1219	4	US-08-935-105A-50	Sequence 50, Appl
35	64	9.1	1219	4	US-09-386-123-50	Sequence 50, Appl
36	64	9.1	1219	4	US-09-648-797-50	Sequence 50, Appl
37	64	9.1	1231	2	US-08-231-193A-48	Sequence 48, Appl
38	64	9.1	1231	2	US-08-486-273A-48	Sequence 48, Appl
39	64	9.1	1231	3	US-08-480-474-48	Sequence 48, Appl
40	64	9.1	1231	3	US-08-940-086A-48	Sequence 48, Appl
41	64	9.1	1231	4	US-08-940-035A-48	Sequence 48, Appl
42	64	9.1	1231	4	US-08-935-105A-48	Sequence 48, Appl
43	64	9.1	1231	4	US-09-648-797-48	Sequence 48, Appl
44	64	9.1	1231	4	US-09-386-123-48	Sequence 48, Appl
45	64	9.1	1235	2	US-08-231-193A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-09-252-991A-17070
; Sequence 17070, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17070
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17070

Alignment Scores:
Pred. No.: 0.861 Length: 539
Score: 75.00 Matches: 25
Percent Similarity: 44.44% Conservative: 3
Best Local Similarity: 39.68% Mismatches: 27
Query Match: 10.68% Indels: 8
DB: 4 Gaps: 3

M64347 (1-385) x US-09-252-991A-17070 (1-539)

Cy 201 AATGCTTCTAGAGTTTATAGCTTGAGCTGCTACCTTTCAAGCTTGGAGGAGCGGTG 260

Db 29 SerAlaAlaArgSerLeuArgSerGlyLeuCysProGlyLysAlaTrpArgProAlaAla 48

QY 261 AATTCAGTGGTTCGTTCTGCTACTGTTACTGGGCGCTGAGTCTGGCGAGCTGCTCCCTTGC 320
Db 49 TrpSerValAlaSer---SerArgGlyThrGlyPro-----AlaAlaCys 62
QY 321 TTGCTCGAGGGCAGTGGCTCAGGGTGGTCTCTCTCTTTGG-----GGCCAGTGCATGCT 374
Db 63 ArgProGlyArgProGlyArgSerValAlaArgTrpSerAlaSerProGlySerPro 82
QY 375 GGCAGAGG 383
Db 83 GlyArgArg 85

RESULT 2
US-08-677-010-3
; Sequence 3, Application US/08677010
; Patent No. 5925805
; GENERAL INFORMATION:
; APPLICANT: Ohlrogge, John B.
; APPLICANT: Roessler, Keith R.
; APPLICANT: Shorosh, Basil S.
; TITLE OF INVENTION: Structure and Expression of an
; TITLE OF INVENTION: Arabidopsis Acetyl-coenzyme A Carboxylase Gene
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,010
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 6550-00002CPA
; TELEPHONE: (810)641-1600
; TELEFAX: (810)641-0270
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2254 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-677-010-3

Alignment Scores:
Pred. No.: 4.65 Length: 2254
Score: 70.50 Matches: 30
Percent Similarity: 38.53% Conservative: 12
Best Local Similarity: 27.52% Mismatches: 40
Query Match: 10.22% Indels: 27
DB: 2 Gaps: 4

M64347 (1-385) x US-08-677-010-3 (1-2254)
QY 364 GGGCCCCAAGAGAGACACCTGAGC-----CATGGCCCTGCAGGC-----AAG 320
Db 1482 GlyProLeuHisGluThrProIleSerAspGlnTyrLysProLeuGlyTyrLeuAspArg 1501
QY 319 CAAGGGACAGTCCGACACTCAGGCCCGCCAGCTACAGTAACAGTACAGAACCAACTGAATTC 260
Db 1502 GlnArgLeuAlaAlaArgSerAsnThrThrTyrCysTyrAspPheProLeuAlaPhe 1521

QY 259 ACGGTTCCCTCCCAAGCTTTGAAGAGTAGCAGT---CCAGGCTATAAACTCTAGAAGCA 203
Db 1522 GlyThrAlaLeuGluLeuLeuTrpAlaSerGlnHisProGlyValLys-----1537
QY 202 TTGCGTAAGAAGTGTTAAGTCTACACAATAATACATCTTTGTAATAAACTCAATAATATAT 143
Db 1538 -----LysProTyr 1540

QY 142 ATATAGATATATATAAACTTTGTAACATCTAATACATCGGAACCTGCACACAGGGCCGGC 83
Db 1541 LysAspThrLeuIleAsnValLysGluLeuValPheSerLysProGluGlySerSerGly 1560

QY 82 CCTCCCTCGGAACCGTCTCCCTGCCT 56
Db 1561 ThrSerLeuAspLeuValGluArgPro 1569

RESULT 3
US-08-790-519-3
; Sequence 3, Application US/08790519
; Patent No. 5962767
; GENERAL INFORMATION:
; APPLICANT: Ohlrogge, John B.
; APPLICANT: Roessler, Keith R.
; APPLICANT: Shorosh, Basil S.
; TITLE OF INVENTION: Structure and Expression of an
; TITLE OF INVENTION: Arabidopsis Acetyl-coenzyme A Carboxylase Gene
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,519
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/248,630
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 6550-00002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810)641-1600
; TELEFAX: (810)641-0270
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2254 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-790-519-3

Alignment Scores:
Pred. No.: 4.65 Length: 2254
Score: 70.50 Matches: 30
Percent Similarity: 38.53% Conservative: 12
Best Local Similarity: 27.52% Mismatches: 40
Query Match: 10.22% Indels: 27
DB: 2 Gaps: 4

M64347 (1-385) x US-08-790-519-3 (1-2254)
QY 364 GGGCCCCAAGAGAGACACCTGAGC-----CATGGCCCTGCAGGC-----AAG 320

Db 1482 GlyProLeuHisGluThrProIleSerAspGlnTyrLysProLeuGlyTyrLeuAspArg 1501
QY 319 CAAGGGAGACGTGCCAGACTCAGGCGCCAGTAAACAGTACAGAACGAAACCACTGAATTC 260
Db 1502 GlnArgLeuAlaAlaArgSerAsnThrThrTyrCysTyrAspPheProLeuAlaPhe 1521
QY 259 ACGCTTCCTCCCAAGCTTTGAAGGTAGAGT---CCAGGCTATAAACTCTAGAAGCA 203
Db 1522 GlyThrAlaLeuGluLeuLeuTyrAlaSerGlnHisProGlyValLys----- 1537
QY 202 TTGGTGAAGAAGTGTAAAGTCTACAAACAATACATCTGTAAACAACTCAATAAATTATAT 143
Db 1538 -----LysProTyr 1540
QY 142 ATATAGATATATAAATTGTAACTTAACTCTAATAACATCGGAACCTCCACACAGGCGCGC 83
Db 1541 LysAspThrLeuIleAsnValLysGluLeuValPheSerLysProGluGlySerSerGly 1560
QY 82 CCTCCCTCGAAACCGTCTCCCTGCCT 56
Db 1561 ThrSerLeuAspLeuValGluArgPro 1569

RESULT 4

US-09-252-991A-16946
; Sequence 16946, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16946
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16946

Alignment Scores:
Pred. No.: 2.76 Length: 205
Score: 70.00 Matches: 31
Percent Similarity: 31.62% Conservative: 12
Best Local Similarity: 22.79% Mismatches: 32
Query Match: 9.97% Indels: 61
DB: 4 Gaps: 7

M64347 (1-385) x US-09-252-991A-16946 (1-205)
QY 1 GACTTCAAGCAAGCTGGTATTTCATAAATCTTCT-----AATGCTGTGTGCC 54
Db 30 AsnThrArgAlaAlaTrpIlePheProProSerSerThrAlaArgCysCys----- 47
QY 55 CAGGCAGGAGACGGTTTCCAGGAGGCGCGCCCTGTGTGCAGTTCGAGTGTATTA 114
Db 48 ---AlaValArgArgTrpAlaAlaSerTrpProAlaSerLys---Pro-Cys----- 63
QY 115 GATGTTACAAGTTTATATATATCTATATATAATTTATTGAGTGTTCACAGATGATT 174
Db 63 ----- 63
QY 175 TGTGTAGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGACTGCTAC 234
Db 64 -----TrpProAlaIle 67
QY 235 CTTTCAAGCTGGAGGAGCGGTGAATTCAGTTGGTTTGTCTGTACTGTACTGGGC 294

Db 67 rValSerProLaAenGlyAlaAlaArgSerAlaCyeTrpIleSerLeuAlaAlaCysTrpAr 87
QY 295 CCTGAGTCTGGCGACTGT-----CCCTTGCTT-----GCCTG 327
Db 87 gArgAlaAlaProLaCysArgArgGlyProIleLeuGlyAlaArgLeuSerProAlaCys 107
QY 328 CAGGCGCAATGGCTCAGGCTGTCTCTTCTTGGGGCCCGCAGTGCATGG 373
Db 107 eArgArgCysSer-----ProSerAlaTrp 115

RESULT 5

US-08-254-989-2
; Sequence 2, Application US/08254989
; Patent No. 5859307
; GENERAL INFORMATION:
; APPLICANT: Mombaerts, Peter
; APPLICANT: Tonegawa, Susumu
; APPLICANT: Johnson, Randall S.
; APPLICANT: Papalocannou, Virginia
; TITLE OF INVENTION: Mutant RAG-1 Deficient Animals Having No
; TITLE OF INVENTION: Mature B and T Lymphocytes
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,989
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/830831
; FILING DATE: 04-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabat, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: MITS783
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 815-6508
; TELEFAX: (404) 815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1040 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mouse
US-08-254-989-2

Alignment Scores:
Pred. No.: 6.66 Length: 1040
Score: 68.50 Matches: 23
Percent Similarity: 44.59% Conservative: 10
Best Local Similarity: 31.08% Mismatches: 22
Query Match: 9.76% Indels: 19
DB: 2 Gaps: 4

M64347 (1-385) x US-08-254-989-2 (1-1040)

QY 84 CCGGCCCTGTGTGCAGGTTCCGATGTTATTAGATGTTACAAGTTTATATATCTATATA 143

Db 464 ProAlaValCysLeu-----AlaIleArgValAsnThrPheLeu----- 476
QY 144 TATAATTTATGAGTTTTCACAGTAT- 173
Db 477 --SerCysSerGlnTyrHisLysMetTyrArgThrValLysAlaIleThrGlyArgGln 495
QY 174 TTGTGTAGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGTGACTGCTA 233
Db 496 IlePheGlnProLeuHisAlaLeuArgAsnAlaGluValLeuLeuProGlyTyrHis 515
QY 234 CCTTTCAAGCTTGGAGGAGCGGTGAATTCAGTTGTTTCG 275
Db 516 PropheGlu---TrpGlnProLeuLysHisValSerSer 528
RESULT 6
US-09-659-166-2
; Sequence 2, Application US/09659166
; Patent No. 6355465
; GENERAL INFORMATION:
; APPLICANT: CREAMY, CARETHA LEE
; APPLICANT: TESTA, TANIA TANSON
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30218
; CURRENT APPLICATION NUMBER: US/09/659,166
; CURRENT FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: UK 9921505.5
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 527
; TYPE: PRT
; ORGANISM: rattus
US-09-659-166-2
Alignment Scores:
Pred. No.: 7.35 Length: 527
Score: 67.50 Matches: 29
Percent Similarity: 42.71% Conservative: 12
Best Local Similarity: 30.21% Mismatches: 39
Query Match: 9.62% Indels: 16
Gaps: 4
DB: 4
M64347 (1-385) x US-09-659-166-2 (1-527)
QY 42 TTGTGTGTGTCGCCAGGAGGAGCGGTTTCCAGGAGGAGGCGGCCCTGTGTGCGAGGT 101
Db 281 LeuLeuLysGlnGlnArgSerSerLeLys-----ValIleAspPheGly 296
QY 102 TCCGATGTTATAGATGTTACAGTTTATATATATATATATATATATATATATATATAT 161
Db 297 SerSer-----CysTyrGluHisGlnArgValTyrThrTyrIleGlnSerArgPhe 313
QY 162 TACAGATGTTATTTGTTGTAGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAG 221
Db 314 Tyr-ArgAlaProGluValIleLeuGlyAlaArgTyrGlyMetProIleAspMetTpsE 333
QY 222 CCGTACTGCTACCTTTCAAGCTTGGAGGAGCGGTGAATTCAGTTGTTGTTCTGT 281
Db 333 rLeuGlyCysIleLeuAlaGluLeu-----LeuThrGlyTyrPr 346
QY 282 ACTGTTACTGGCGCTGTGGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 327
Db 346 oLeuLeuProGlyGluAspGluGlyAsp---GlnLeuAlaCysMet 360
RESULT 7
US-08-802-466-2
; Sequence 2, Application US/08802466
; Patent No. 5972606
; GENERAL INFORMATION:
; APPLICANT: Creamy, et al.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,466
; FILING DATE: 19 February 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GH50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 528 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-802-466-2
Alignment Scores:
Pred. No.: 8.49 Length: 528
Score: 67.00 Matches: 23
Percent Similarity: 45.07% Conservative: 9
Best Local Similarity: 32.39% Mismatches: 30
Query Match: 9.54% Indels: 9
Gaps: 2
DB: 2
M64347 (1-385) x US-08-802-466-2 (1-528)
QY 117 TCTTCAAGTTTAT 176
Db 300 CysTyrGluHisGlnArgValTyrThrTyrIleGlnSerArgPheTyr-ArgAlaProGln 319
QY 177 TTGTAGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGTGACTGCTACT 236
Db 319 uValIleLeuGlyAlaArgTyrGlyMetProIleAspMetTpsSerLeuGlyCysIleLe 339
QY 237 TTCAAAGCTTGGAGGAGCGGTGAATTCAGTTGTTGTTCTGTCTGTCTGTCTGTCTGT 296
Db 339 uAlaGluLeu-----LeuThrGlyTyrProLeuLeuProGlyGln 352
QY 297 TGAGTCTGGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 327
Db 352 uAspGluGlyAsp---GlnLeuAlaCysMet 361
RESULT 8
US-09-350-484-2
; Sequence 2, Application US/09350484
; Patent No. 6159716
; GENERAL INFORMATION:
; APPLICANT: Creamy, et al.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road

```

CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION NUMBER: 08/802,466
APPLICANT: FastSeq for Windows Version 2.0
FILING DATE: 09-Jul-1999
CLASSIFICATION: <Unknown>
PRIORITY: 08/802,466
ATTORNEY/AGENT INFORMATION:
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: GH50002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 528 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-350-484-2

Alignment Scores:
Pred. No.: 8.49 Length: 528
Score: 67.00 Matches: 23
Percent Similarity: 45.07% Conservative: 9
Best Local Similarity: 32.39% Mismatches: 30
Query Match: 9.54% Indels: 9
DB: Gaps: 2

M64347 (1-385) x US-09-350-484-2 (1-528)
QY 117 TGTACAAGTTTATATATATATATATATATATATATATATATATATGATTTTAAAGATGATTTTG 176
Db 300 CysTyrGluHisGlnArgValTyrThrIleGlnSerArgPheTyr-ArgAlaProGI 319
QY 177 TTGTAGACTTACACTTCTTAGCGCAATGCTTCTAGAGTTTATAGCTTGACTGCTACCT 236
Db 319 uValIleLeuGlyAlaArgTyrGlyMetProIleAspMetTrpSerLeuGlyCysIleLe 339
QY 237 TTCAAAGCTTGGAGGGAAGCCCGTGAATTCAGTTTGTTCGTCTGTACTGTACTTGGGCC 296
Db 339 uAlaGluLeu-----LeuThrGlyTyrProLeuLeuProGlyGI 352
QY 297 TGAGTCTGGGCACTGTGCTCTGTTCGCTG 327
Db 352 uAspGluGlyAsp---GlnLeuAlaCysMet 361

```

```

; RESULT 9
; US-09-122-126B-15
; Sequence 15, Application US/091222:126B
; Patent No. 6451575
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES
; FILE REFERENCE: DM6909
; CURRENT APPLICATION NUMBER: US/09/122,126B
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15

```

```

; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-122-126B-15

Alignment Scores:
Pred. No.:      9.93      Length:      930
Score:          67.00     Matches:      38
Percent Similarity: 30.22% Conservative: 17
Best Local Similarity: 20.88% Mismatches:   54
Query Match:       9.71% Indels:         73
DB:                4      Gaps:          7

M64347 (1-385) x US-09-122-126B-15 (1-930)

Qy    379 TGGCCACATGCACT----- 365
      ||| |||||
Db    452 TrpSerLysCysThrSerAlaThrIleThrGluPheLeuAspAspGlyHisGlyAsnCys 471
      ::::
Qy    364 -----CGGCCCAAGAA----- 353
      ::::
Db    472 LeuLeuAspLeuProArgLysGlnIleLeuGlyProGluGluLeuProGlyGlnThrTyrr 491
      ::::
Qy    352 -----GAGACCACCTCGAGCATGCCT-----GCAGGCAAGCAA 317
      ::::
Db    492 AspAlaThrGlnGlnCysAsnLeuThrPheGlyProGluTyrrSerValCysProglyMet 511
      ::::
Qy    316 GGGACAGCTGCCAGACTCAGGGCCCAGTACAGTACAGACGAACAACCACTCAATTCAAG 257
      |||:|||:|||||
Db    512 AspValCysAlaArgLeuTrpCysAlaValValargGlnGlyGlnMetValCysLeuThr 531
      |||:|||:|||||
Qy    256 GTTTCCCTCCAAGCTTTGAAAGGTAGCAGTCCTAGGCTAT----- 218
      |||:|||:|||||
Db    532 LysLysLeuProAlaValGluGlyThrProCysGlyLysGlyArgIleCysLeuGlnGly 551
      |||:|||:|||||
Qy    217 AAAAECTTAGAAGCATTCGCTAAGAAGTTTAAGTCTACACA----- 175
      |||:|||:|||||
Db    552 LysCysValAspLysThrLysLysLys-TyrTyrrSerThrSerSerHisGlyAsnTrpGl 571
      |||:|||:|||||
Qy    174 -----AATACATCTTTGAAAAAACTCAATAAAATTATATATA 140
      ::::|||||
Db    571 ySerTrpGlySerTrpGlyGlnCysSerArgSerCysGlyGlyGlyValGlnPheAlaTy 591
      ::::|||||
Qy    139 TAGATATATATAAATTGTAACTCTTAATAACACTCGGAACCTGCACACAGGCGCCGCCCC 80
      |||:|||:|||||
Db    591 rArGHis-----CysAsn-----AsnProAlaPr 599
      ::::|||||
Qy    79 TCCCTGGAAACCGCTCCCTCCCTGGGACACACAGCAATTAGAAGAATTTGTATCAAAAT 20
      |||:|||:|||||
Db    599 oArgAsnAsnGlyArgTyrrCysThrGlyLysArgAlaIleTyrrargSerCysSerLeuMe 619
      ::::|||||
Qy    19 ACCA 16
      :|||
Db    619 tPro 620

```

```

RESULT 10
US-09-634-286A-15
; Sequence 15, Application US/09634286A
; Patent No. 6521436
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGREGAN DGRADING METALLO PROTEASES
; FILE REFERENCE: DM6909A
; CURRENT APPLICATION NUMBER: US/09/634,286A
; CURRENT FILING DATE: 2000-08-09
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-634-286A-15

```

```

Alignment Scores:
Pred. No.: 9.93 Length: 930
Score: 67.00 Matches: 38
Percent Similarity: 30.22% Conservative: 17
Best Local Similarity: 20.88% Mismatches: 54
Query Match: 9.71% Indels: 73
DB: 4 Gaps: 7

M64347 (1-385) x US-09-634-286A-15 (1-930)
QY 379 TGGCCACCATGCACT----- 365
DB 452 TrpSerLysCysThrSerAlaThrIleThrGluPheLeuAspGlyHisGlyAsnCys 471
QY 364 -----GGGCCCCAAGAA----- 353
DB 472 LeuLeuAspLeuProArgLysGlnIleLeuGlyProGluGluLeuProGlyGlnThrTyr 491
QY 352 -----GAGACCACCTGAGCCATGGCCCT-----GCAGGCAAGCAA 317
DB 492 AspAlaThrGlnGlnCysAsnLeuThrPheGlyProGluTyrSerValCysProGlyMet 511
QY 316 GGGACAGCTGCCAGACTCAGGGCCCACTAACAGTACAGAACGAACTGAATTCACG 257
DB 512 AspValCysAlaArgLeuTyrCysAlaValValArgGlnGlyGlnMetValCysLeuThr 531
QY 256 GCTTCCCTCCCAAGCTTTGAAGGTAGCAGTCCAGGCTAT----- 218
DB 532 LysLysLeuProAlaValGluGlyThrProCysGlyLysGlyArgIleCysLeuGlnGly 551
QY 217 AAAAATCTAGAACGATTCGTAAGTGTAAAGTCTACAAACA----- 175
DB 552 LysCysValAspLysThrLysLysLys-TyrTyrSerThrSerSerHisGlyAsnTrpG1 571
QY 174 -----AATACATCTGTAAAAAATCAATAAATATATATA 140
DB 571 ySerTrpGlySerTrpGlyGlnCysSerArgSerCysGlyGlyValGlnPheAlaTy 591
QY 139 TAGATATATATAAATTTGTAACTCTAATACATCGGAACCTGCACACAGGGCGGCCCC 80
DB 591 rArgHis-----CysAsn-----AsnProAlaPr 599
QY 79 TCCCTGGAACCCGTCCTCCCTCGCTGGGACACAGCAATTAGAAGATTGTATGAAAT 20
DB 599 cArgAsnAsnGlyArgTyrCysThrGlyLysArgAlaIleTyrArgSerCysSerLeuMe 619
QY 19 ACCA 16
DB 619 tPro 620

RESULT 11
US-08-185-432-18
; Sequence 18, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Busseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DSLTDX PROTEINS, NUCLEIC ACIDS, AND
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

Alignment Scores:
Pred. No.: 15.1 Length: 2523
Score: 66.50 Matches: 20
Percent Similarity: 39.06% Conservative: 5
Best Local Similarity: 31.25% Mismatches: 26
Query Match: 9.64% Indels: 13
DB: 1 Gaps: 3

M64347 (1-385) x US-08-185-432-18 (1-2523)
QY 201 TGCCTAAGAGTGTAGTCTACAAACAATACATCTTGTAAAACTCA-----ATA 151
DB 176 CysLysGlnAspIleAsnGluCysSerGlnAsnProCysLysAsnGlyGlnCysIle 195
QY 150 AATATATATATAGATATATATAAATCTGTAACTCTTAATACATCGGAACCTGCACACA 91
DB 196 AsnGluPheGlySerTyrArgCysThrCys-----GlnAsnArgPheThr 210
QY 90 GGG-----CCGGCCCTCCCTGGAAACCGTCTCCCTCGCTGGGACACACA 46
DB 211 GlyArgAsnCysAspGluProTyrValProCysAsnProSerProCysLeuAsnGlyGly 230
QY 45 GCAATTAGAAGA 34
DB 231 ThrCysArgGln 234

RESULT 12
US-08-899-232-3
; Sequence 3, Application US/08899232
; Patent No. 6436650
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Oii, HuiLin
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
; FILE REFERENCE: 7326-046
; CURRENT APPLICATION NUMBER: US/08/899,232
; CURRENT FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2523
; TYPE: PRT
; ORGANISM: Xenopus sp.
US-08-899-232-3

Alignment Scores:
Pred. No.: 15.1 Length: 2523
Score: 66.50 Matches: 20
Percent Similarity: 39.06% Conservative: 5
Best Local Similarity: 31.25% Mismatches: 26
Query Match: 9.64% Indels: 13
DB: 1 Gaps: 3

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2523 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-185-432-18

```

DB: 4 Gaps: 3

M64347 (1-385) x US-08-999-232-3 (1-2523)

Qy 201 TGCCTAAGAGTGTTAAAGTGTACAAACAATACATCTTGTAAAAACTCA-----ATA 151
||| :|: :|: ||| ||| |||
Db 176 CysLysGlnAspIleAsnGluCysSerGlnAsnProCysLysAsnGlyGlnCysIle 195
||| :|: :|: ||| ||| |||
Qy 150 AATTATATATAGATATATATAAACCTTGTAACATCTTAATAACATCGGAACCTGGCACACA 91
||| :|: :|: ||| ||| |||
Db 196 AsnGluPheGlySerTyArgCysThrCys-----GlnAsnArgPheThr 210
||| :|: :|: ||| ||| |||
Qy 90 GGG-----CCGGCCCCCTCCCTGGAAACCGCTCCCTGCCCTGGGACACACA 46
||| :|: :|: ||| ||| |||
Db 211 GlyArgAsnCysAspGluProTyValProCysAsnProSerProCysLeuAsnGlyGly 230
||| :|: :|: ||| ||| |||
Qy 45 GCAATTAGAACA 34
||| :|: :|: ||| ||| |||
Db 231 ThrCysArgGln 234
||| :|: :|: ||| ||| |||

RESULT 13

US-09-252-991A-30876
; Sequence 30876, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; LENGTH: 140
; TYPE: PR1
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30876

Alignment Scores:

Pred. No.:	Score:	Length:
303	65.50	140
Percent Similarity:	44.90%	Matches: 17
Best Local Similarity:	34.69%	Conservative: 5
Query Match:	9.49%	Mismatches: 14
DE:	4	Indels: 13
		Gaps: 3

M64347 (1-385) x US-09-252-991A-30876 (1-140)

Qy 344 CCTGAGCATGGCCCTTCAGGCAAGCA-----AGGGACAGTGCACA 303
||| :|: :|: ||| ||| |||
Db 80 ProSerProTrpProCysArgProAlaProValSerArgArgTyArgProAlaMetPro 99
||| :|: :|: ||| ||| |||
Qy 302 GAC-----TCAGGCCCGCATACAGTACAGAACGAACCAACTGAATTCAGGCTCCCT 249
||| :|: :|: ||| ||| |||
Db 100 SerSerProSerAlaProProThrSerSerArgArgCysSer-----Pro 114
||| :|: :|: ||| ||| |||
Qy 248 CCAAGCTTTGAAGTAGCAGTCCAGG 222
||| :|: :|: ||| ||| |||
Db 115 ProGlyTrpArgArgGlySerAlaArg 123
||| :|: :|: ||| ||| |||

RESULT 14

US-09-252-991A-21105
; Sequence 21105, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

```
QY 352 -----GAGACCACTGAGCCATGGCCCT-----GCAGGCAAGCAA 317
Db   :|||:|
492 AspAlaThrGlnGlnCysAsnLeuThrPheGlyProGluTyrSerValCysProGlyMet 511
QY 316 GGGACAGCTGCCGAGACTCAGGGCCAGTAACAGTACAGAACGAAACCACTGAATTCACG 257
Db   :|||:|
512 AspValCysAlaArgLeuTrpCysAlaValValArgGlnGlnMetValCysLeuThr 531
QY 256 GCTTCCCTCCAAAGCTTCAAGGTAGCAGTCCAGCTAT----- 218
Db   :|||:|
532 LysLysLeuProAlaValGluGlyThrProCysGlyLysGlyArgValCysLeuGlnGly 551
QY 217 AAACTCTAGAAGCATTCGCTAAGAGTGTTAAGTCTACAACA----- 175
Db   :|||:|
552 LysCysValAspLysThrLysLysLys-TyrTyrSerThrSerHisGlyAsnTrpG1 571
QY 174 -----AATACATCTGTAAAACTCAATAAATTATATATA 140
Db   :|||:|
571 ySerTrpGlyProTrpGlyGlnCysSerArgSerCysGlyGlyGlyValGlnPheAlaTy 591
QY 139 TAGATATATATAAACTTGTAAACATCTAATAACATCGGAACCTGCACACAGGCGCGCCCC 80
Db   :|||:|
591 rArgHis-----CysAsn-----AsnProAlaPr 599
QY 79 TCCCTGGAAACGGTCTCCCTCGCTGGGACACACAGCAATTAGAAGAATTTGTATGAAAAT 20
Db   :|||:|
599 oArgAsnSerGlyArgTyrCysThrGlyLysArgAlaIleTyrArgSerCysSerValTh 619
QY 19 ACCA 16
Db   :|||
619 rPro 620
```

Search completed: September 22, 2004, 10:56:32
Job time : 30.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 22, 2004, 10:50:56 ; Search time 54.5 Seconds
(without alignments)
4537.112 Million cell updates/sec

Title: M64347

Perfect score: 702

Sequence: 1 GACTCAAGCAAGCTGGTA.....GTGCATGGTGGCCAGAGGTG 385

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 2684796

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV-xlp

-Q=/cgn2_1/USFTO.spool_p/CANELLA305B/runat_22092004_112338_14585/app_query.fasta_1.583
-DB=Published Applications_AA -QMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NCPU=200000000 -USER=CANELLA305B @CGN 1 1 225 @runat 22092004_112338_14585
-NCPUL=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOC=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubpaa/CTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	74.5	10.6	357	16	US-10-767-701-47126	Sequence 47126, A
2	74	10.5	87	14	US-10-029-386-28395	Sequence 28395, A
3	72.5	10.3	147	15	US-10-264-237-2482	Sequence 2482, Ap
4	71	10.1	1045	15	US-10-161-927-36	Sequence 36, Appl
5	71	10.1	2854	16	US-10-085-198-16	Sequence 16, Appl
6	70.5	10.0	1355	16	US-10-437-963-124522	Sequence 124522,
7	70.5	10.2	1385	16	US-10-437-963-123966	Sequence 123966,
8	70	10.0	1106	16	US-10-437-963-196290	Sequence 124520,
9	69.5	9.9	672	16	US-10-437-963-124521	Sequence 124521,
10	69.5	10.1	2214	15	US-10-369-493-6737	Sequence 6737, Ap
11	69	10.0	569	12	US-10-087-192-207	Sequence 207, App
12	68.5	9.8	135	12	US-10-424-599-158314	Sequence 158314,
13	68.5	9.8	391	14	US-10-017-161-2324	Sequence 2324, Ap
14	68.5	9.8	391	15	US-10-292-798-1970	Sequence 1970, Ap
15	68	9.6	168	14	US-10-157-031-228	Sequence 228, App
16	67.5	9.6	116	16	US-10-437-963-198137	Sequence 198137,
17	67.5	9.8	118	12	US-10-424-599-203789	Sequence 203789,
18	67.5	9.6	141	9	US-09-925-299-787	Sequence 787, App
19	67.5	9.6	141	10	US-09-925-299-787	Sequence 787, App
20	67.5	9.8	396	14	US-10-198-070-94	Sequence 94, Appl
21	67.5	9.6	526	9	US-09-855-154-2	Sequence 2, Appl
22	67.5	9.8	599	14	US-10-279-733-9	Sequence 9, Appl
23	67	9.5	525	16	US-10-437-963-177003	Sequence 177003,
24	67	9.5	528	14	US-10-253-349-6	Sequence 6, Appl
25	67	9.7	628	12	US-10-358-283-8	Sequence 8, Appl
26	67	9.7	752	12	US-10-358-283-6	Sequence 6, Appl
27	67	9.7	753	12	US-10-358-283-4	Sequence 4, Appl
28	67	9.7	763	12	US-10-358-283-15	Sequence 15, Appl
29	67	9.7	930	12	US-10-358-283-2	Sequence 2, Appl
30	67	9.7	930	12	US-10-358-283-30	Sequence 30, Appl
31	67	9.7	930	14	US-10-247-685-15	Sequence 15, Appl
32	67	9.7	1359	16	US-10-408-765A-218	Sequence 218, App
33	67	9.7	1411	13	US-10-408-765A-237	Sequence 237, App
34	67	9.5	1638	13	US-10-090-458-2	Sequence 2, Appl
35	66.5	9.5	124	16	US-10-437-963-191569	Sequence 191569,
36	66.5	9.5	147	12	US-10-424-599-165824	Sequence 165824,
37	66.5	9.5	259	15	US-10-369-493-20007	Sequence 20007, A
38	66.5	9.6	450	16	US-10-437-963-149314	Sequence 149314,
39	66.5	9.6	491	16	US-10-437-963-157538	Sequence 157538,
40	66.5	9.6	1043	9	US-09-946-805-4	Sequence 4, Appl
41	66.5	9.6	2524	15	US-10-190-115-25	Sequence 25, Appl
42	66.5	9.6	2524	15	US-10-369-072-25	Sequence 25, Appl
43	66	9.4	161	14	US-10-156-761-12304	Sequence 12304, A
44	66	9.4	193	16	US-10-437-963-194777	Sequence 194777,
45	66	9.4	294	12	US-10-343-650A-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1

US-10-767-701-47126
; Sequence 47126, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 47126
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(357)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C7832_1.pep

Publication No. US20030235821A1
GENERAL INFORMATION:
APPLICANT: Zerhusen, Bryan D.
APPLICANT: Kekuda, Ramesh
APPLICANT: Spytek, Kimberly A.
APPLICANT: Shenoy, Suresh G.
APPLICANT: Miller, Charles E.
APPLICANT: Hjal, Tord
APPLICANT: Gerlach, Valerie L.
APPLICANT: Baumgartner, Jason C.
APPLICANT: Guo, Xiaojia
APPLICANT: Gangolli, Esha A.
APPLICANT: Vernet, Corine
APPLICANT: Padigar, Muralidhara
APPLICANT: Li, Li
APPLICANT: Pena, Carol E.A.
APPLICANT: Gorman, Linda
APPLICANT: Anderson, David W.
APPLICANT: Eisinger, Shlomit R.
APPLICANT: Patturajan, Meera
APPLICANT: Stone, David J.
TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
FILE REFERENCE: 21402-377 D (Cura 677 Other)
CURRENT APPLICATION NUMBER: US/10/161,927
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: 60/295,661
PRIOR FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 60/295,607
PRIOR FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 60/296,404
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: 60/296,418
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: 60/296,575
PRIOR FILING DATE: 2001-06-07
PRIOR APPLICATION NUMBER: 60/297,414
PRIOR FILING DATE: 2001-06-11
PRIOR APPLICATION NUMBER: 60/297,567
PRIOR FILING DATE: 2001-06-12
PRIOR APPLICATION NUMBER: 60/298,528
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 60/325,685
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/299,133
PRIOR FILING DATE: 2001-06-18
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 190
SEQ ID NO 36
LENGTH: 1045
TYPE: PRT
ORGANISM: Homo sapiens
US-10-161-927-36
Alignment Scores:
Pred. No.: 46.7 Length: 1045
Score: 71.00 Matches: 24
Percent Similarity: 40.24% Conservative: 9
Best Local Similarity: 29.27% Mismatches: 25
Query Match: 24 Indels: 24
DB: 5 Gaps: 5
M64347 (1-385) x US-10-161-927-36 (1-1045)
QY 189 CACTTCTACGCAATGCTTCTAGAGTTTATAGCCTGGACTGCTACCT-----TTCAA 242
Db 225 HsArgLeuHisAnsSerValSerAlaProSerProGlyVilProProArgaspPheLys 244
QY 243 GCT-----TGGAGGGAAGCCCGTGAATCA 266
Db 245 SerLeuAlaLeuAlaArgAlaProGlyHisGlyGlyPheTrp---GlnGlyValAlaAla 263
QY 267 GTTGGTTCGTTCTGTACTGTACTGGG-----CCCTGAGTCTGGGCGAGCTGTC 314

Db 264 GluGlyValGlyCysThrLeuThrGlyAlaTrpArgSerProValProTrpSerGlyThr 283
QY 315 CCTTGCTTGCTGCA-----GGGCGATGGCTCAGGGTGGTCTCTCTCTGG 359
Db 284 GlyCysValProGlyGlyPheThrValProGlyProArgProAlaProAlaProTrp 303
QY 360 GGCCCA 365
Db 304 GlyPro 305
RESULT 5
US-10-085-198-16
Sequence 16, Application US/10085198
Publication No. US2004000907A1
GENERAL INFORMATION:
APPLICANT: Alsobrook et al.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-279
CURRENT APPLICATION NUMBER: US/10/085,198
CURRENT FILING DATE: 2002-02-25
PRIOR APPLICATION NUMBER: 60/271,646
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 60/276,401
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/311,981
PRIOR FILING DATE: 2001-08-13
PRIOR APPLICATION NUMBER: 60/312,858
PRIOR FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: 60/271,840
PRIOR FILING DATE: 2001-02-27
PRIOR APPLICATION NUMBER: 60/277,324
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: 60/286,096
PRIOR FILING DATE: 2001-04-21
PRIOR APPLICATION NUMBER: 60/299,695
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/315,614
PRIOR FILING DATE: 2001-08-29
PRIOR APPLICATION NUMBER: 60/272,405
PRIOR FILING DATE: 2001-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 653
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 16
LENGTH: 2854
TYPE: PRT
ORGANISM: Homo sapiens
US-10-085-198-16

Alignment Scores:
Pred. No.: 55.6 Length: 2854
Score: 71.00 Matches: 25
Percent Similarity: 43.06% Conservative: 6
Best Local Similarity: 34.72% Mismatches: 25
Query Match: 10.11% Indels: 16
DB: 15 Gaps: 3
M64347 (1-385) x US-10-085-198-16 (1-2854)
QY 182 GACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCCTGGACTGCTACCTTTCAA 241
Db 819 GluLeuThrLeuLeu-----TrpAspArgThr----- 827
QY 242 AGCTTGGAGGAAGCCGGAATTCAGTTGGTTCCTGTACTGTCTTACTGGCC----- 295
Db 828 GlyValProGlyGlySerGluLeuSerPhePheLeuGluProTrpArgSerSer 847
QY 296 ---CTGAGTCTGGGAGCTGCTCCCTTGGTTCCTGCGAGGGCCATGGCTCAGGGTGGTCTC 352
Db 848 CysThrSerTrpSerSerCys--LeuGlyCysLeuAlaAspGlnGlyCysGlyTrpCysL 867
QY 353 TTCITGGGGCCCGCATGTCATGTCGCCAGGT 384


```

; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 207
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-087-192-207

Alignment Scores:
Pred. No.: 71.5 Length: 569
Score: 69.00 Matches: 40
Percent Similarity: 35.80% Conservative: 18
Best Local Similarity: 24.69% Mismatches: 46
Query Match: 10.00% Indels: 58
DB: 12 Gaps: 10

M64347 (1-385) x US-10-087-192-207 (1-569)
QY 380 CTGGCCACCATGCACTGGGCCCCCAAGAGACACCACTGAGCATGGCC----- 330
DB 60 LeuAlaArgGlnTyrTrpLeuPheAlaAlaArgAlaHisProGlnPro---ProArgAlaAsp 78
QY 329 -----TCAGCGCAAGCAAGGAC 312
DB 79 LeuValSerLeuGlnPheAlaGluLeuPheGlnArgHisPheCysArgGluValArgGlu 98
QY 311 AGCTGCCAGACTCAGGSCCCAGTAACAGTACAGAACCACTGAATTCACGGCTTC 252
DB 99 SerLeu-----AlaGlyProProGlyHisAspTyrArgAlaThr-----Ala 112
QY 251 CCTCCAGCTTTGAAGAGTCCAGCTCCAGGCTATAAACTCTAGAACCATTCGCTAAGAA 192
DB 113 ProPro-----ArgProAlaLeuProLysAlaArgSerSer-----Glu 125
QY 191 GTGTGA-----AGTCTACAACAATACTATCTTGTAATAAACTCAATAA 150
DB 126 AspLeuGlyProArgProAlaCysAlaLeuGlnHisLeuArgArgGlyLeuArgGln--- 144
QY 149 ATTATATATAGATATATATAA-----CTTGTA 120
DB 145 -----LeuPheArgArgSerAlaGlyGluLeuProGlyAlaThrSer 159
QY 119 ACATCTAATACATCGGAACCTGCACAGGCGCGCCCTCCCTCGAAACCGCTCCCT 60
DB 160 AspThrAsnAspIleAspThrThrAlaAlaSerArgProGlyProAlaArgLysLeuLeu 179
QY 59 GCCTGG-----GACACACAGCAATTAGAAGAAATTTGTATGAATAACACAG 15
DB 180 ProTrpGlyLeuArgGluProProThrGluAlaLeuLysGluValValLeuArgTyrSer 199
QY 14 CTGTCT 9
DB 200 LeuAla 201

RESULT 12
US-10-424-599-158314
; Sequence 158314, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yinhua
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 158314
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_113977C.1.pep
US-10-424-599-158314

Alignment Scores:
Pred. No.: 63.5 Length: 135
Score: 68.50 Matches: 24
Percent Similarity: 38.95% Conservative: 6
Best Local Similarity: 31.17% Mismatches: 20
Query Match: 9.76% Indels: 27
DB: 12 Gaps: 4

M64347 (1-385) x US-10-424-599-158314 (1-135)
QY 206 TTCTAGAGTTTTATAGCCTGGACTGCTACCTTCAAGCTTGGAGGGAAGCCGTGAATTC 265
DB 55 PheValSerValValSerPheGlyAlaThr-----Val 65
QY 266 AGTTGGTTGCTCTCTACTACTGTTAC-----TGGGCCCTGAGTCTGGGAGCTGT 313
DB 66 SerGluPheValValMetCysValValPheGlyTyrPheLeuIleTyrGlySerCys 85
QY 314 CCCTGCTGCTCCCTGGAGGCCATGGCT-----CAGGGTGGTCTCTCTTTG----- 358
DB 86 ProLeuThrLeuCysAspLeuSerAlaSerGlyArgArgGlyGlyCysPheLeuGluHis 105
QY 359 -----GGGCCAGTGCATGTCGCCACAG 382
DB 106 LysValPheGluIleHisMetCysAlaGlyGlySerIleGlyTyrGlnGln 122

RESULT 13
US-10-017-161-2324
; Sequence 2324, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAL, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2324
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (213)
; OTHER INFORMATION: Variable amino acid
US-10-017-161-2324

Alignment Scores:
Pred. No.: 76.5 Length: 391
Score: 68.50 Matches: 29

```

Percent Similarity: 40.00% Conservative: 11
 Best Local Similarity: 29.00% Mismatches: 32
 Query Match: 9.76% Indels: 28
 DB: 14 Gaps: 6

M64347 (1-385) x US-10-017-161-2324 (1-391)

```

QY 73 CCAGGGAGGGCGCGCCCTGTGTCAGGTTCCGATGTTATAGATGTTACAGTTTATAT 132
Db 196 ProPheArgSerArgSerCys-----ProLeuLeuSerValProSerAlaLeu 211
QY 133 ATATCTATATATATAATTTATTAGATGTTTACAGATGTTTGT----- 180
Db 212 Val***Phe-----ValSerCysArgPheValCysCysLeuSerIleAla 226
QY 181 -----AGACTTAACACTTCTTACGCAATGCTTGTAGATGTTTATAGCTGACTCTAC 234
Db 227 SerGlyArgValArgSer-CysValSerCysPhe---SerPheValArgPheValArgLe 245
QY 235 CTTTCAAAGCTTGGAGGAGCCGCGTGAATTCAGTTGGTTCTGTTCTGTAC----- 283
Db 245 uPhe-----ArgPheIleSerTrpPheLeuArgPheArgSerSerVa 259
QY 284 ----TGTTACTGGCCCTGAGTCTGGCAGCTGTCCTTGTCTGCTGCGGCGCATG 337
Db 259 lSerSerPheValArgLeuSerLeuGlyLeuCysSerCysLeuValLeuArgSerVal 278

```

RESULT 14

US-10-292-798-1970

```

; Sequence 1970, Application US/10292798
; Publication No. US2003023583A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1970
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (213)..(213)
; OTHER INFORMATION: Variable amino acid
US-10-292-798-1970

```

Alignment Scores:

Pred. No.: 76.5 Length: 391
 Score: 68.50 Matches: 29
 Percent Similarity: 40.00% Conservative: 11
 Best Local Similarity: 29.00% Mismatches: 32
 Query Match: 9.76% Indels: 28
 DB: 15 Gaps: 6

M64347 (1-385) x US-10-292-798-1970 (1-391)

```

QY 73 CCAGGGAGGGCGCGCCCTGTGTCAGGTTCCGATGTTATAGATGTTACAGTTTATAT 132
Db 196 ProPheArgSerArgSerCys-----ProLeuLeuSerValProSerAlaLeu 211
QY 133 ATATCTATATATATAATTTATTAGATGTTTACAGATGTTTGT----- 180
Db 212 Val***Phe-----ValSerCysArgPheValCysCysLeuSerIleAla 226

```

```

QY 181 -----AGACTTAACACTTCTTACGCAATGCTTGTAGATGTTTATAGCTGACTCTAC 234
Db 227 SerGlyArgValArgSer-CysValSerCysPhe---SerPheValArgPheValArgLe 245
QY 235 CTTTCAAAGCTTGGAGGAGCCGCGTGAATTCAGTTGGTTCTGTTCTGTAC----- 283
Db 245 uPhe-----ArgPheIleSerTrpPheLeuArgPheArgSerSerVa 259
QY 284 ----TGTTACTGGCCCTGAGTCTGGCAGCTGTCCTTGTCTGCTGCGGCGCATG 337
Db 259 lSerSerPheValArgLeuSerLeuGlyLeuCysSerCysLeuValLeuArgSerVal 278

```

RESULT 15

US-10-157-031-228

```

; Sequence 228, Application US/10157031
; Publication No. US20030108990A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 228
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (163)..(163)
; OTHER INFORMATION: X = unknown
US-10-157-031-228

```

Alignment Scores:

Pred. No.: 75.4 Length: 168
 Score: 68.00 Matches: 19
 Percent Similarity: 53.33% Conservative: 5
 Best Local Similarity: 42.22% Mismatches: 12
 Query Match: 9.86% Indels: 9
 DB: 14 Gaps: 2

M64347 (1-385) x US-10-157-031-228 (1-168)

```

QY 356 AGAAGAGACCACCTGAGCCATGCCCTGCAGCAAGGACAGCTGCCAGACTCA 297
Db 80 ArgArgThrArgProGluPro-GlyProAla---ArgArgGlyAlaArgAlaGluProPr 98
QY 296 GGGCCAGTAAACAGTACAGAACCAACTGAATTCACGGCTTCCTCCCAAGCTTTGAA 237
Db 98 oGlyProProThrLeuGlnAspProProAlaSerHisLeu----- 111
QY 236 AGTAGAGAGTCCA 224
Db 112 -GlySerSerPro 115

```

Search completed: September 22, 2004, 10:58:25
 Job time : 59.5 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 22, 2004, 10:45:25 ; Search time 30 seconds
(without alignments)
2468.916 Million cell updates/sec

Title: M64347

Perfect score: 702

Sequence: 1 GACTTCAAAGCAAGCTGTA.....GTGCATGTGCCAGAGGTG 385

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cpn2.1/USPTO.spool_p/CANEIL3A305B/runat_22092004.112336.14518/app.query.fasta_1.583
-DB=PIR_78 -OFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODES=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=CANEIL3A305B @CGN 1 1 63 @runat_22092004.112336.14518 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	85.5	12.4	674	2 A97490	hypothetical prote
2	74.5	10.6	263	2 T15453	hypothetical prote
3	73.5	10.5	1073	2 I51055	recombination acti
4	73	10.4	591	2 T39195	probable amino aci
5	72.5	10.3	404	2 T19922	hypothetical prote
C 6	71	10.3	187	2 S35050	muslin JER28 - huma
7	71	10.1	324	2 C90492	hypothetical prote
8	70.5	10.0	360	2 C81699	conserved hypothet
9	70.5	10.0	1042	2 S42511	RAG-1 protein - ra
10	70.5	10.0	1043	2 B33754	recombination-acti
C 11	70.5	10.2	2257	2 D86483	protein F3J5.19 li
C 12	69.5	10.1	2214	2 T16305	hypothetical prote
13	69	9.8	2090	2 T30075	hypothetical prote
14	69	9.8	2153	2 T30074	hypothetical prote

C 15	68.5	9.9	605	2 G95853	probable pyruvate
16	68	9.7	296	2 S21306	hypothetical prote
17	68	9.7	849	2 T39406	hypothetical prote
18	67.5	9.6	363	2 G71544	hypothetical prote
C 19	67	9.7	326	2 A46676	CD68 homolog macro
20	67	9.5	326	2 H72472	hypothetical prote
C 21	66.5	9.6	1043	2 D84900	hypothetical prote
C 22	66.5	9.6	2524	2 A35844	Xotch protein - Af
C 23	66	9.6	606	2 S43118	finger protein - m
24	65.5	9.3	335	2 T25054	hypothetical prote
C 25	65.5	9.5	863	2 T47038	hypothetical prote
C 26	65.5	9.5	863	2 AD0234	probable fibrillar
C 27	65	9.3	361	2 D84452	hypothetical prote
C 28	64.5	9.3	166	2 D83230	hypothetical prote
29	64.5	9.2	255	2 T41451	very hypothethical
30	64.5	9.2	590	1 S34960	NADH2 dehydrogenas
31	64.5	9.2	603	2 S06059	gene ND1 intron 4
32	64.5	9.2	645	2 T25824	hypothetical prote
C 33	64	9.3	235	2 T31424	C-terminal domain-
C 34	64	9.3	348	2 H70549	probable pdhs prot
C 35	64	9.3	3429	2 T13853	hypothetical prote
36	63.5	9.0	918	2 I58178	glutamate receptor
37	63.5	9.0	949	2 S19808	glutamate receptor
C 38	63.5	9.2	1615	2 JE0372	low density lipopr
39	63	9.0	275	2 T20246	hypothetical prote
40	63	9.0	412	2 F86296	hypothetical prote
C 41	63	9.1	454	2 T35380	probable membrane
C 42	63	9.0	474	2 D91041	probable PTS enzym
43	63	9.0	474	2 G85885	probable PTS enzym
C 44	63	9.1	1970	2 T03284	myoblast city prot
C 45	62.5	9.1	125	2 S20639	19 heavy chain V r

ALIGNMENTS

RESULT 1

A97490
Hypothetical protein AGR_C_1968 [imported] - Agrobacterium tumefaciens (strain C58, Cerc
C/Species: Agrobacterium tumefaciens
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
R/Accession: A97490
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A/Reference number: A97959; MUID:21608551; PMID:11743194
A/Accession: A97490
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-674 <KUR>
A/Cross-references: GB:AE007869; PIDN:AKR86874.1; PID:GL5156092; GSPDB:GN00169
C/Genetics:
A/Gene: AGR_C_1968
A/Map position: circular chromosome

Alignment Scores:

Pred. No.: 0.119 Length: 674
Score: 85.50 Matches: 27
Percent Similarity: 43.59% Conservative: 7
Best Local Similarity: 34.62% Mismatches: 25
Query Match: 12.39% Indels: 19
DB: 2 Gaps: 3

M64347 (1-395) x A97490 (1-674)

QY 376 CCACATGCTACTGGGCCCCAGAA-----GAG 350
|||||
Db 310 ProProAlaLeuGlnProGlnGlnAlaThrProSerGlnProSerAlaAlaThrSerGlu 329
349 ACCACCTGACCCATGGCCCTGCAGCAAGCAGGACAGTGCACAGCTCAG----- 296
||||| : : : : :
Db 330 ThrThrAlaAsnThrProValProSerSerGlyThrAlaAlaProThrAlaAlaGly 349

QY 295 -----GCCAGTAACA-----GTACAGAACGACCAACTGAATTCACGGCT 254
 Db 350 SerThrAlaValAlaProLeuThrSerSerValGlnValThrAlaValGluPheGluGly 369
 QY 253 TCCCTCCAAAGCTTTGAAGGTAGCAGTCCAGGCTATAAACTCTAGAAGCATTTG 200
 Db 370 SerLysIlePheValAlaGlySerAlaProGlyGlySerThrValArgAlaLeu 387
 RESULT 2
 T15453
 hypothetical protein CO8A9.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
 C:Accession: T15453
 R:Latreille, P.
 submitted to the EMBL Data Library, December 1995
 A:Description: The sequence of C. elegans cosmid CO8A9.
 A:Reference number: Z18353
 A:Accession: T15453
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-263 <LAT>
 A:Cross-references: EMBL:U42844; NID:g1125797; PID:g1125799; PIDN:AAB53817.1; GSPDB:GN00
 A:Experimental source: strain Bristol N2; clone CO8A9
 C:Genetics:
 A:Gene: CESP:CO8A9.3
 A:Map position: X
 A:Introns: 45/3; 79/3; 128/2
 Alignment Scores:
 Pred. No.: 2.57 Length: 263
 Score: 74.50 Matches: 29
 Percent Similarity: 39.29% Conservatives: 15
 Best Local Similarity: 25.89% Mismatches: 25
 Query Match: 10.61% Indels: 43
 DB: 2 Gaps: 5
 M64347 (1-385) x T15453 (1-263)
 QY 28 ACAAATCTCTTAATTCGTGTGCCAGGAGGAGCGTTCCAGGAGGAGGCGCG 87
 Db 140 SerAsnSerAlaIleIleCysLeu-----GlyArgGlyGln 151
 QY 88 CCCTGTGCGAGGTTCCGATGTTATTAGATGTACAAAGTTTATATATATATATA 147
 Db 152 -----ileGlyIleTyrLeu 156
 QY 146 ATTTATTGAGTTTACAGATGT----- 171
 Db 157 ValTyrTyrAlaValGlnLysCysArgPheGluArgGlnSerPheThrLeuPheTyrLys 176
 QY 172 ATTTGTTGTAGACTTAACACTTCTTAGCAATGCTTCTAGAGTTTATAGCCTGGACTGC 231
 Db 177 IleCysCysThrLeuIlePheIleValPheMetLeuMetGlu-----CysLeuAsnArg 194
 QY 232 TACCTTCAAGCTTGGAGGAGCCGTGAATTCAGTTGCTTCTGTTACTGTTACTG 291
 Db 195 TyrLeuAlaAsn-----PheMetLeuThrTyrAsnValLeuLeuThr 208
 QY 292 GGCCCTGAGTCTGGCAGCTGCCCTGCTGCTGCTG 327
 Db 209 ProAlaLysSerArgGlnLeuAsnMetAspCysVal 220
 RESULT 3
 I51055
 recombinant activating protein - rainbow trout
 C:Species: Oncorhynchus mykiss (rainbow trout)
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: I51055
 R:Hansen, J.D.; Kaattari, S.J.
 Immunogenetics 42, 188-195, 1995
 A:Title: The recombination activation gene 1 (RAG1) of rainbow trout (Oncorhynchus mykiss)
 A:Reference number: I51055; MUID:95369845; PMID:7642230

A:Accession: I51055
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1073 <HAN>
 A:Cross-references: EMBL:U15663; NID:9558917; PIDN:AAA80281.1; PID:9558918
 C:Genetics:
 A:Gene: RAG-1
 A:Introns: 489/1
 C:Superfamily: RING finger homology
 C:Keywords: zinc
 F:306-354/Domain: RING finger homology <RNG>
 Alignment Scores:
 Pred. No.: 3.17 Length: 1073
 Score: 73.50 Matches: 28
 Percent Similarity: 44.83% Conservatives: 11
 Best Local Similarity: 32.18% Mismatches: 28
 Query Match: 10.47% Indels: 20
 DB: 2 Gaps: 4
 M64347 (1-385) x I51055 (1-1073)
 QY 59 CAGGAGACGGTTTCCAGGAGGCGCGCTGTGTCAGGTTCCGATGTTATGATG 118
 Db 488 GlnGlyArgGlyPhe-GlyLeuHisProAlaValCysLeu-----AlaIleArgVa 504
 QY 119 TTCAAGTTTAT 173
 Db 504 LAsnThrPheLeu-----SerCysSerGlnTyrHisLysMetTyrArgTh 519
 QY 174 -----TTGTTGTAGACTTAACACTTCTTACGAATGCTTC 208
 Db 519 rValLysAlaThrSerGlyArgGlnIlePheGlnProLeuHisThrLeuArgThrAlaGl 539
 QY 209 TAGAGTTTATAGCTGAGCTGCTACCTTCAAAAGCTTGGAGGAGCCGCTGAATTCAGT 268
 Db 539 ulysGluLeuLeuProGlyTyrHisProPheGlu---TrpGlnProAlaLeuLysSerVa 558
 QY 269 TGGTTCGTTCTGTACTGTT 287
 Db 558 lSerThrSerCysHisVal 564
 RESULT 4
 T39195
 probable amino acid permease - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: T39195
 R:Wedler, H.; Duesterhoeft, A.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, October 1999
 A:Reference number: Z21834
 A:Accession: T39195
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-591 <WED>
 A:Cross-references: EMBL:AL121764; PIDN:CA57428.1; GSPDB:GN000666; SPDB:SPAC9.10
 A:Experimental source: strain 972h-; cosmid C9
 C:Genetics:
 A:Gene: SPDB:SPAC9.10
 A:Map position: 1
 C:Superfamily: choline transport protein
 Alignment Scores:
 Pred. No.: 3.74 Length: 591
 Score: 73.00 Matches: 33
 Percent Similarity: 39.34% Conservatives: 15
 Best Local Similarity: 27.05% Mismatches: 28
 Query Match: 10.40% Indels: 46
 DB: 2 Gaps: 9
 M64347 (1-385) x T39195 (1-591)
 QY 149 TTTATTGAGTTTTTACAAGATGATTTGTTGTAGAC----- 184

```

Db      45 PheGlnGluGlyPheGluAspValSerValThrAspAspAspAsnAsnGluLeuLeu 64
QY      185 -----TTAACACTTCTACGCAATGCTTCTAGAGTTTATAGCTGGACT 229
Db      65 ArgLysMetGlyTyrGlnProValLeuHisArgSerPheGluPheGluSerPheAla 84
QY      230 GTACCTCTCAAGAGCTGGAG-----GGAAGCCGT---GAATTCAAGTTGGTTCGTT 277
Db      85 AlaSerPheAlaSerLeuAspValSerGlyValArgLeuThrPheSerTrpGlyIle 104
QY      278 CTGTAC-----TGTTACTGGCCCTGAGTCTG-----GGCAGCTGCCCTTG 319
Db      105 SerPheGlyGlyProAlaAlaTyrTrpSerAlaMetLeuValThrGlyPheCysSerIle 124
QY      320 CTGTCC-----TGCAGGGCCCATG---GCTCAGGGTGGTCTCTTC 355
Db      125 ValThrAlaAlaCysLeuAlaGluIleCysSerAlaLeuProAlaAlaGlySerIleTyr 144
QY      356 TTG-----GGGCCAGT-----GCA 370
Db      145 LeuTrpAlaAlaGluSerAlaGlyProArgPheGlyArgPheValSerPheLeuValAla 164
QY      371 TGGTGG 376
Db      165 TrpTrp 166

RESULT 5
T19922
hypothetical protein C44C10.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T19922
R:Ottage, A.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19197
A:Accession: T19922
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-404 <WIL>
A:Cross-references: EMBL:Z69787; PIDN:CAA93635.1; GSPDB:GN00028; CESP:C44C10.3
A:Experimental source: clone C44C10
C:Genetics:
A:Gene: CESP:C44C10.3
A:Map position: X
A:Introns: 73/2; 158/3; 290/3; 376/3

Alignment Scores:
Pred. No.: 4.37 Length: 404
Score: 72.50 Matches: 31
Percent Similarity: 35.88% Conservative: 16
Best Local Similarity: 23.66% Mismatches: 29
Query Match: 10.33% Indels: 55
DB: 2 Gaps: 7

M64347 (1-385) x T19922 (1-404)
QY      11 CAAGCTGGTATTTTCATACAAATCTCTTAATGCTGTG-----TGTCCAGGCAG 61
Db      271 GlnSerGlyMetIleLeuSerPheLeuIleLeuAlaIlePheLeuTrpIleAspTyrHis 290
QY      62 GGAGACGGTTTCAGGAGGGCCGCTGTGTGTCAGTTCGATGTTTAGATGTTA 121
Db      291 GlyThrGlyPhe-----Leu 295
QY      122 CAAGTTTATATATATATATATATATATATATATTTATTAGTGTTTTCAAGATGATTTGTGTA 181
Db      296 IleValTyrLeuPheGlyThrMet-----PheIleGluTyrThrTrpAspAla----- 311
QY      182 GACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGGACTGCTACCTTCAA 241
Db      312 -----ValTyrLeuCysAlaIleGlu 318

```

```

QY      242 AGTTTGAGGAGGAGCCGCTGAATTCAGTTGTTGTTCTGTACTGTACTGGCCCTCAGT 301
Db      319 SerMetGluThrSerSerArg-----AlaSerAla 328
QY      302 CTGGGCACTGCTCCCTTTCCTGCTGCTGAGGCGCATGCTCAGGTCGTCTCTTCTTGGGG 361
Db      329 valGlySerCysSerLeu-----MetAlaArgIleGlySerLeuLeuAla 343
QY      362 CCC-----AGTGCATGTGGCCA 379
Db      344 PropheLeuThrTyrAlaAsnThrTrpTrpPro 354

RESULT 6
S35050
mucin JER28 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 10-Dec-1993 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: S35050; S37593
R:Dufosse, J.; Porchet, N.; Audie, J.P.; Guyonnet Duperat, V.; Laine, A.; van-Seuningen,
Biochem. J. 293, 329-337, 1993
A:Title: Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic alternati
A:Reference number: S35047; MUID:93343858; PMID:7916618
A:Accession: S35050
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-187 <DUF>
A:Experimental source: clone JER28
A:Note: this publication is not cited in GenBank entry HSMUC5BB, release 113.0
R:Aubert, J.
submitted to the EMBL Data Library, September 1993
A:Reference number: S37593
A:Accession: S37593
A:Molecule type: mRNA
A:Residues: 'S', 2-11, 'E', 13-143, 'N', 145-187 <AUB>
A:Cross-references: EMBL:X74954; NID:9407069; PIDN:CAA52909.1; FID:9407070

Alignment Scores:
Pred. No.: 6.85 Length: 187
Score: 71.00 Matches: 25
Percent Similarity: 45.31% Conservative: 4
Best Local Similarity: 39.06% Mismatches: 25
Query Match: 10.29% Indels: 10
DB: 2 Gaps: 4

M64347 (1-385) x S35050 (1-187)
QY      367 ACTGGGCCCCAAGAGAGACCCCTGAGCCATGCGCTGAGGCAAGCAAGGACAGCT 308
Db      11 ThrGlyProSerThrThrAlaThrValThr---GlyProThrGlySerThrAlaThrAla 29
QY      307 GCCCAGACTCAGGGCCCCAGTA-----ACAGTACAGAACGACCAACT 266
Db      30 SerSerThrGlnAlaThrAlaGlyThrProHisValSerThrThrAlaThrThrProThr 49
QY      265 GAATTCAGGCTTCCTCCCAAGCTTTGAAGGTAGGTCAGGCTATAAACT---CTA 209
Db      50 -----ValThrSerSerLysAlaThrProPheSerSerProGlyThrAlaThrAlaLeu 67
QY      208 GAAGCATTTGCGT 197
Db      68 ProAlaLeuArg 71

RESULT 7
C90492
hypothetical protein SSO3088 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: C90492
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.

```

A;Reference number: A99139
A;Accession: C90492
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-324 <KUR>
A;Cross-references: GB:AE01
C;Genetics:
A;Gene: SS03088

Alignment Scores:		
Pred. No.:	6.68	Length:
Score:	71.00	Matches:
Percent Similarity:	41.30%	Conservative:
Best Local Similarity:	28.26%	Mismatches:
Query Match:	10.11%	Indels:
DB:	2	Gaps:

M64347 (1-385) x C90492 (1-324)

1	GACTTCAAGCAAGCTGGTATATTTTCATACAAATCTCTTAATTCGTCTGTGTCCTCCACGCA	60
253	AspPheLys---AsnTrpArgPheHisCysGluLysAspIleCysLeuValSerSerAla	271
61	GGGAGACCGTTTCCA---GGAGGGGGCCCGCTGTGTGCAGGTTCCGATGTATTATGAT	117
272	-----ProLeuGlyArgGlyTrp-----	277
118	GTTACAAGTTTATATATATATATATATATATATATTTATTGAGTTTTTACA	165
278	-----LeuPheThrGlyLeuIleLys	284

166 QY -----AGATGATTATTGTTAGACCTTAACACTTCTTACGCCAAGCTCTTAGAGAGTTTAT 219
 285 Db SerAsnLysMetValcysAlaValcysAlaGlyIleAspThrGluAsnGluThrLeuLeuAspTyrGln 304
 220 QY AGCCTGGACTGCTACCTTTCAAGCTTGGAGGGAAG 255
 305 Db AlaLeuGlu---TyrLeuLysLysLeuAlaGlyGlu 315

RESULT 8
C81699
conserved hypothetical protein TC0468 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 18-Aug-2000
C:Accession: C81699
R:Read, T.D.; Brnham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: C81699
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-360 <TET>
A:Cross-references: GB:AE002315; GB:AE002160; NID:G7190506; PIDN:AAF39317.1; PID:G7190514
A:Experimental source: strain Nigg (MoPn)

Alignment Scores:	
Pred. No.:	7.63
Score:	70.50
Percent Similarity:	34.93%
Best Local Similarity:	21.92%
Query Match:	10.04%
DR:	2
	Length:
	Matches:
	Conservative:
	Mismatches:
	Indels:
	Gaps:
	360
	32
	28
	67
	7

M64347 (1-385) x C81699 (1-360)

Qy 91 TGTGTGAGGTTCCGATGTTATTAGATGTTACAAGTTTATAT----- 132
 ||||: :||:||||: :||| |||||
 Db 178 CysileuLeuproilellethrValArqValLeuTvrrAsnAlpheArqPheLeu 197

```

A:Residues: 1-2257 <STO>
A:Cross-references: GB:AE005172; NID:gl2039052; PIDN:AAF18638.2; GSPDB:GN001414
C:Genetics:
A:Gene: F5J5.19
A:Map position: 1
C:Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl

```

```

Alignment Scores:
Pred. NO.: 7.01 Length: 2257
Score: 70.50 Matches: 30
Percent Similarity: 38.53% Conservative: 12
Best Local Similarity: 27.52% Mismatches: 40
Query Match: 10.22% Indels: 27
DB: 2 Gaps: 4

```

```

M64347 (1-385) x D86483 (1-2257)
QY 364 GGGCCCCAAGAGAGACACCCTGAGC-----CATGCCCTGCAGGC-----AAG 320
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1485 GlyProLeuHisGluThrProIleSerAspGlnTyrLysProLeuGlyTyrLeuAspArg 1504
QY 319 CAAGGGACAGCTGCCACAGACTCAGGCCCCAGTAACAGTACAGACGAACCAACTGAAATTC 260
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1505 GlnArgLeuAlaAlaArgSerAsnThrThrTyrCysTyrAspPheProLeuAlaPhe 1524
QY 259 ACGGCTTCCTCCCAAGCTTTGAAGGTAGCAGT---CCAGGCTATAAACTCTAGAACCA 203
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1525 GlyThrAlaLeuGluLeuLeuThrAlaSerGlnHisProGlyValLys----- 1544
QY 202 TTGCGTAGAAGTGTTAAGTCTACACAAATACATCTTGTAATAAACTCAATAAATTATAT 143
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1541 -----LysProTyr 1544
QY 142 ATATAGATATATATAAACTTTGAACATCTAATACATCGGAACCTGCACACAGGCCCGGC 83
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1544 LysAspThrLeuIleAsnValLysGluLeuValPheSerLysProGluGlySerSerGly 1563
QY 82 CCCTCCCTCGAAACCGTCTCCCTGCCT 56
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1564 ThrSerLeuAspLeuValGluArgPro 1572

```

```

RESULT 12
T16305
hypothetical protein F40F4.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16305
R:Wilson, R.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F40F4.
A:Reference number: Z18493
A:Accession: T16305

```

```

A:Molecule type: DNA
A:Residues: 1-2214 <WIL>
A:Cross-references: EMBL:UA0420; NID:G1065513; PID:G1065514; PIDN:AAA81430.1;
C:Genetics:
A:Gene: CESP.F40P4.6
A:Introns: 57/3; 95/1; 302/3; 323/1; 386/1; 404/1; 468/1; 507/2; 547/3; 595/3;

Alignment Scores:
Pred. NO.: 9.25 Length: 2214
Score: 69.50 Matches: 21
Percent Similarity: 48.53% Conservative: 12
Best Local Similarity: 30.88% Mismatches: 24
Query Match: 10.07% Indels: 11
DB: 2 Gaps: 4

M64347 (1-385) x T16305 (1-2214)
QY 373 CCATGACTGGGCCCCNAGNAGAGACCCCTGAGCATGGCCCTGCAGGCAAGCAA--- 317
Db 1369 ProlleThrMetProThrGlnGlnThrAlaLeuThrGlnGlyProValThrGlnGlnThr 198

```


A; Experimental source: strain 1021, megaplasmid pSymb
 R; Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Barloy-Hubler,
 Peia, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A; Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A; Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A; Reference number: A96039; MUID:21368234; PMID:11474104
 A; Contents: annotation
 C; Genetics:
 A; Gene: SMB20095
 A; Genome: plasmid
 C; Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain B

Alignment Scores:
 Pred. No.: 12.9 Length: 605
 Score: 68.50 Matches: 23
 Percent Similarity: 33.02% Conservative: 12
 Best Local Similarity: 21.70% Mismatches: 36
 Query Match: 9.93% Indels: 35
 DB: 2 Gaps: 3

M64347 (1-385) x G95853 (1-605)

QY	325	GGCAAGCAGGACAGCTGCCAGACTCAGGGCCCACTAACAGTACAG-----	278
Db	518	GlyIleMetAspGlnAlaPheAlaValAspGlyProValIleGlnAlaLeuValAsp	537
QY	277	-----AACGAACCAACTGAATTCACGGCTTCCCTCCAAGCT	242
Db	538	ArgTyrGluProLeuMetProProLysMetProAlaAspTyrAlaArgAsnPheArgAla	557
QY	241	TTGAAAGGTAGCTCAGCTATAAACTCTAGAGCATTCGTAAGAGTGTAAAGTC	182
Db	558	AlaLeuProGluThrProGlyHisGluLysIleGluLileuLeuArgAsnSerSerVal	577
QY	181	TACAACAATACATCTTGTAAAACTCAATAATTATATATATATATATAAATTG	122
Db	578	GlyArgLys-----	580
QY	121	TAACATCTATAATACATCGGAA-----CCTGCACACAGGGCGGCCCTCCCTGGAA	71
Db	581	-----ValThrAspGluGluProGlnProHisGluAlaAlaProAspGluAsn	597
QY	70	ACCGTCTCCCTGCTGGG	53
Db	598	ThrGlyGluLeuProgly	603

Search completed: September 22, 2004, 10:55:40
 Job time : 38 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 22, 2004, 10:37:10 ; Search time 17.5 Seconds
(without alignments)
2291.087 Million cell updates/sec

Title: M64347

Perfect score: 702

Sequence: 1 GACTTCAAGCAAGCTGGTA.....GTGCATGTGGCCAGAGGTG 385

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool.p/CANELLA305B/runat_22092004_112335_14494/app_query.fasta_1.583
-DB=SwissProt_42 -QFWT=fastan -SUFFIX=resp -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cgi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=CANELLA305B @cgn_1.1.33 @runat_22092004_112335_14494 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73.5	10.5	1073	RAG1_ONCMY	Q91187 oncorhynch
2	73	10.4	591	YFYA_SCHPO	Q9u18 schizosacch
3	70.5	10.0	1040	RAG1_MOUSE	P15919 mus musculu
4	70.5	10.0	1042	RAG1_RABIT	P34088 oryctolagus
5	69	10.0	548	LNK_MOUSE	O09039 mus musculu
6	67.5	9.8	484	IRX5_MOUSE	Q9jkg4 mus musculu
7	67	9.7	211	VNS3_RSVN	Q01210 rice stripe
8	67	9.7	211	VNS3_RSVT	P26658 rice stripe
9	67	9.7	326	CD68_MOUSE	P31996 mus musculu
10	67	9.5	528	DYR2_HUMAN	Q92630 homo sapien
11	67	9.7	930	AT55_HUMAN	Q9una0 homo sapien
12	67	9.7	1411	TCOF_HUMAN	Q13428 homo sapien
13	66.5	9.6	2524	NOTC_XENLA	P21783 xenopus lae
14	65.5	9.3	523	AAAL_HUMAN	Q9ns82 homo sapien
15	65	9.3	351	U204_ARATH	Q925q3 arabidopsis
16	65	9.4	930	AT55_MOUSE	Q9r001 mus musculu
17	64.5	9.2	353	VGLM_MCMVK	P52373 murine cyto
18	64.5	9.3	833	CASL_MOUSE	Q35177 mus musculu

19	64.5	9.2	1233	1	NME3_HUMAN	Q14957 homo sapien
20	64	9.1	228	1	COB5_PYRFU	Q8u400 pyrococcus
21	64	9.3	378	1	LFNG_MOUSE	Q09010 mus musculu
22	64	9.3	378	1	LFNG_RAT	Q924t4 rattus norv
23	63.5	9.2	297	1	HXAB_CHICK	P31258 gallus gall
24	63.5	9.0	918	1	GLK1_HUMAN	P39086 homo sapien
25	63.5	9.0	949	1	GLK1_RAT	P22756 rattus norv
26	63	9.1	365	1	LXCL_HUMAN	Q9nzus homo sapien
27	63	9.1	518	1	TEX5_HUMAN	Q99593 homo sapien
28	62.5	9.1	240	1	F171_HALNI	Q7hs17 halobacteri
29	62.5	9.1	310	1	FMT_GLOVI	Q7hjk1 gloeobacteri
30	62.5	8.9	430	1	DGOT_ECOLI	P31457 escherichia
31	62.5	9.1	732	1	POK_DROME	Q01842 drosophila
32	62	8.8	301	1	HVGL_PSEXP	P13703 pseudomonas
33	62	8.8	357	1	YMR2_EBY	P03192 epstein-bar
34	62	8.8	474	1	YFEV_ECOLI	P77272 escherichia
35	62	9.0	1146	1	AGE1_CABEL	Q94125 caenorhabdi
36	62	8.8	1305	1	CYA9_XENLA	P98999 xenopus lae
37	62	8.8	7073	1	RIAB_CVHSA	P59641 h replicase
38	61.5	8.8	399	1	OYE2_YEAST	Q03558 saccharomyc
39	61.5	8.9	414	1	IF2G_HALNI	Q9hmk9 halobacteri
40	61.5	8.8	530	1	AAAL_MOUSE	Q9imh8 mus musculu
41	61.5	8.8	554	1	YCT3_SCHPO	Q5913 schizosacch
42	61.5	8.8	559	1	DACA_HUMAN	Q9Byj9 homo sapien
43	61.5	8.8	875	1	NETR_HUMAN	P56730 homo sapien
44	61.5	8.8	940	1	PTGA_MYCPN	P75569 mycoplasma
45	61.5	8.8	1043	1	RAG1_HUMAN	P15918 homo sapien

ALIGNMENTS

RESULT 1
RAG1_ONCMY STANDARD; PRT; 1073 AA.
AC Q91187.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE V(D)J recombination activating protein 1 (RAG-1).
GN RAG1 OR RAG-1.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Shaeta; TISSUE=nestis;
RX MEDLINE=93369845; PubMed=7642230;
RA Hansen J.D., Kaattari S.I.;
RT "The recombination activation gene 1 (RAG1) of rainbow trout
RT (Oncorhynchus mykiss): cloning, expression, and phylogenetic
RT analysis.";
RL Immunogenetics 42:188-195(1995).
CC !- FUNCTION: During lymphocyte development, the genes encoding
CC immunoglobulins and T cell receptors are assembled from variable
CC (V), diversity (D), and joining (J) gene segments. This
CC combinatorial process, known as V(D)J recombination, allows the
CC generation of an enormous range of binding specificities from a
CC limited amount of genetic information. The RAG1/RAG2 complex
CC initiates this process by binding to the conserved recombination
CC signal sequences (RSS) and introducing a double-strand break
CC between the RSS and the adjacent coding segment. These breaks are
CC generated in two steps, nicking of one strand (hydrolysis),
CC followed by hairpin formation (transesterification). RAG1/2 has
CC also been shown to function as a transposase in vitro, and to
CC possess RSS-independent endonuclease activity (end processing) and
CC hairpin opening. RAG1 alone can bind to RSS but stable, efficient
CC binding requires RAG2. All known catalytic activities require the
CC presence of both proteins (By similarity).
CC !- COFACTOR: Binds 1 magnesium or manganese ion per subunit (By
CC similarity).
CC !- SUBCELLULAR LOCATION: Nuclear.

CC -!- TISSUE SPECIFICITY: Thymus, a lower level expression is seen in
 CC the kidney.
 CC -!- DOMAIN: The specific binding to the nonamer RSS motif is mediated
 CC by the nonamer binding domain (NBD) (By similarity).
 CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
 CC -!- SIMILARITY: Contains 1 nonamer binding (NBD) domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U15663; AA80281.1; -
 CC PIR; I51055; I51055.
 CC HSSP; P15919; LRMD.
 CC InterPro: IPR001841; Znf ring.
 CC Pfam; PF00097; zf-C3HC4_1.
 CC SMART; SM00184; RING_1.
 CC PROSITE; PS00518; ZF_RING_1; 1.
 CC PROSITE; PS00089; ZF_RING_2; 1.
 CC Hydrolase; EC:3.1.1.1.
 CC DNA recombination; Zinc-finger; Metal-binding.
 CC FT ZN FING 310 349 RING-TYPE.
 CC FT DN5 BIND 421 476 NBD.
 CC FT METAL 533 633 DIVALENT METAL ION (BY SIMILARITY).
 CC FT METAL 743 743 DIVALENT METAL ION (BY SIMILARITY).
 CC FT ACT SITE 597 597 BY SIMILARITY.
 CC SQ SEQUENCE 1073 AA; 121044 MW; C08930808B709B5 CRC64;
 CC
 CC Alignment Scores:
 CC Pred. No.: 2,12 Length: 1073
 CC Score: 73.50 Matches: 28
 CC Percent Similarity: 44.93% Conservative: 11
 CC Best Local Similarity: 32.18% Mismatches: 28
 CC Query Match: 10.47% Indels: 20
 CC DB: 1 Gaps: 4
 CC
 CC M64347 (1-385) x RAG1_ONCMY (1-1073)
 CC QY 59 CAGGAGACGGTTTCACGGAGGGCGCGCTGTGTCAGGTTCCGATGTTATTAGATG 118
 CC Db 488 GlnGlyAeGlyPhe-GlyLeuHisProAlaValCysLeu-----AlaIleArgVa 504
 CC QY 119 TTACAAGTTTATATATATCTATATATATAATTTTGTGAGTTTTCACAGATGTAT----- 173
 CC Db 504 lAsnThrPheLeu-----SerCysSerGlnTyrHisLysMetTyrArgTh 519
 CC QY 174 -----TTGTTGTAGACTTAACACTTCTTACGCAATGCTTC 208
 CC Db 519 rVallysAlaThrSerGlyAeGlnIlePheGlnProLeuHisThrLeuArgThrAlaGl 539
 CC QY 209 TAGAGTTTATAGCTGCTGCTACCTTCAAGCTTCGAGGAGCGCGGAATTCAGT 268
 CC Db 539 ulysGluLeuLeuProGlyTyrHisProPheGlu---TrpGlnProAlaLeuLysSerVa 558
 CC QY 269 TGGTTCGTCTGTACTGT 287
 CC Db 558 lSerThrSerCysHisVal 564

RESULT 2

YFYA_SCHPO STANDARD; PRT; 591 AA.
 AC QUT18;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative amino-acid permease C9.10.
 GN SPAC9.10.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 CC NCB1_TaxID=4896;
 CC [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=972;
 CC RX MEDLINE=21848401; PubMed=11859360;
 CC RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 CC Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 CC Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 CC Collins M., Connor R., Cronin A., Davis P., Fiddall J., Fraser A.,
 CC Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 CC Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 CC James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 CC Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 CC Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 CC Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 CC Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 CC Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 CC Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 CC Welter J., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
 CC Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 CC Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 CC Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 CC Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 CC Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 CC Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 CC Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 CC Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 CC Cerutti L., Lowe T., McConbie M.R., Paulsen I., Potashkin J.,
 CC Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 CC RA The genome sequence of Schizosaccharomycetes pombe.";
 CC RL Nature 415:871-880(2002).
 CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: Belongs to the amino acid permease family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL121764; CAB57428.1; -
 CC PIR; T39195; T39195.
 CC GeneDB; SPombe; SPAC9.10; -
 CC InterPro; IPR002293; AA/rel permease1.
 CC InterPro; IPR004840; AAC permease.
 CC InterPro; IPR004841; Permease region.
 CC Pfam; PF00324; aa_permeases; 1.
 CC PROSITE; PS00218; AMINO-ACID-PERMEASE_1; FALSE_NEG.
 CC KW Hypothetical protein; Transport; Amino-acid transport;
 CC Transmembrane.
 CC FT TRANSMEM 98 118 POTENTIAL.
 CC FT TRANSMEM 342 362 POTENTIAL.
 CC FT TRANSMEM 397 417 POTENTIAL.
 CC FT TRANSMEM 450 470 POTENTIAL.
 CC FT TRANSMEM 545 565 POTENTIAL.
 CC SQ SEQUENCE 591 AA; 65069 MW; FFCD707AC7849D4 CRC64;
 CC
 CC Alignment Scores:
 CC Pred. No.: 2,28 Length: 591
 CC Score: 73.00 Matches: 33
 CC Percent Similarity: 39.34% Conservative: 15
 CC Best Local Similarity: 27.05% Mismatches: 28
 CC Query Match: 10.40% Indels: 46
 CC DB: 1 Gaps: 9
 CC
 CC M64347 (1-385) x YFYA_SCHPO (1-591)
 CC QY 149 TTTATTGAGTTTTCACAGATGTTTGTGTAGAC----- 184

```

Db      45 PheGlnGluGlyPheGluAspValSerValThrAspAspAsnAspAsnGluLeuLeu 64
QY      185 -----TTACACCTTTCAGCAATGCTTCTAGAGTTTATACGCTGGACT 229
Db      65 ArgLysMetGlyTyrGlnProValLeuHisArgSerPheGluPheGluSerPheAla 84
QY      230 GCTACCTTTCAAGCTTGGAG-----GGAGACCGT---GAATTCAGTTGGTTTCGTT 277
Db      85 AlaserPheAlaserLeuAspValValSerGlyValArgLeuThrPheSerTrpGlyIle 104
QY      278 CTGTAC-----TGTTACTGGCCCTGAGTCTG-----GGCAGCTGTCCCTTG 319
Db      105 SerPheGlyGlyProAlaAlaTyrTTPSerAlaMetLeuValThrGlyPheCysSerIle 124
QY      320 CTGTGCC-----TCGAGGCCCATG---GCTCAGGGTGGTCTCTTC 355
Db      125 ValThrAlaAlaCysLeuAlaGluIleCysSerAlaLeuProAlaGlySerIleTyr 144
QY      356 TTG-----GGGCCCATG-----GCA 370
Db      145 LeuTrpAlaAlaGluSerAlaGlyProArgPheGlyArgPheValSerPheLeuValAla 164
QY      371 TGTGG 376
Db      165 TrpTrp 166

RESULT 3
RAG1_MOUSE
ID      RAG1_MOUSE      STANDARD;      PRT;      1040 AA.
AC      P15919;
DT      01-APR-1990 (Rel. 14, Created)
DT      01-APR-1990 (Rel. 14, Last sequence update)
DE      28-FEB-2003 (Rel. 41, Last annotation update)
DE      V(D)J recombination activating protein 1 (RAG-1).
GN      RAG1 OR RAG-1.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90030604; PubMed=2598259;
RA      Schatz D.G., Oettinger M.A., Baltimore D.;
RT      "The V(D)J recombination activating gene, RAG-1."
RL      Cell 59:1035-1048(1989).
RN      [2]
RP      MUTAGENESIS.
RX      MEDLINE=94111984; PubMed=8284210;
RA      Sadofsky M.J., Hesse J.E., McElane J.F., Gellert M.;
RT      "Expression and V(D)J recombination activity of mutated RAG-1
RT      proteins."
RL      Nucleic Acids Res. 21:5644-5650(1993).
RN      [3]
RP      MUTAGENESIS.
RX      MEDLINE=20069477; PubMed=10601032;
RA      Landree M.A., Wibbenmeyer J.A., Roth D.B.;
RT      "Mutational analysis of RAG1 and RAG2 identifies three catalytic amino
RT      acids in RAG1 critical for both cleavage steps of V(D)J
RT      recombination."
RL      Genes Dev. 13:3059-3069(1999).
RN      [4]
RP      MUTAGENESIS.
RX      MEDLINE=20069478; PubMed=10601033;
RA      Kim D.R., Dai Y., Mundy C.L., Yang W., Oettinger M.A.;
RT      "Mutations of acidic residues in RAG1 define the active site of the
RT      V(D)J recombinase."
RL      Genes Dev. 13:3070-3080(1999).
RN      [5]
RP      MUTAGENESIS OF ASP-600 AND ASP-708.
RX      MEDLINE=20142662; PubMed=10678172;
RA      Fugmann S.D., Valley I.J., Praszek L.M., Schatz D.G.;
RT      "Identification of two catalytic residues in RAG1 that define a single
RT      active site within the RAG1/RAG2 protein complex."

```

```

Mol. Cell 5:97-107(2000).
[6]
X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 265-380.
MEDLINE=97372897; PubMed=9228952;
RA      Bellon S.F., Rodgers K.K., Schatz D.G., Coleman J.E., Steitz T.A.;
RT      "Crystal structure of the RAG1 dimerization domain reveals multiple
RT      zinc-binding motifs including a novel zinc binuclear cluster."
RL      Nat. Struct. Biol. 4:586-591(1997).
CC      !- FUNCTION: During lymphocyte development, the genes encoding
CC      immunoglobulins and T cell receptors are assembled from variable
CC      (V), diversity (D), and joining (J) gene segments. This
CC      combinatorial process, known as V(D)J recombination, allows the
CC      generation of an enormous range of binding specificities from a
CC      limited amount of genetic information. The RAG1/RAG2 complex
CC      initiates this process by binding to the conserved recombination
CC      signal sequences (RSS) and introducing a double-strand break
CC      between the RSS and the adjacent coding segment. These breaks are
CC      generated in two steps, nicking of one strand (hydrolysis),
CC      followed by hairpin formation (transesterification). RAG1/2 has
CC      also been shown to function as a transposase in vitro, and to
CC      possess RSS-independent endonuclease activity (end processing) and
CC      hairpin opening. RAG1 alone can bind to RSS but stable, efficient
CC      binding requires RAG2. All known catalytic activities require the
CC      presence of both proteins.
CC      !- COFACTOR: Binds 1 magnesium or manganese ion per subunit.
CC      !- SUBCELLULAR LOCATION: Nuclear.
CC      !- TISSUE SPECIFICITY: MATURING LYMPHOID CELLS AND CENTRAL NERVOUS
CC      SYSTEM.
CC      !- DOMAIN: The specific binding to the nonamer RSS motif is mediated
CC      by the nonamer binding domain (NBD) (By similarity).
CC      !- SIMILARITY: Contains 1 RING-type zinc finger.
CC      !- SIMILARITY: Contains 1 nonamer binding (NBD) domain.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
EMBL; M29475; AAA40028.1; .
PDB; 1BMD; 23-JUL-97.
MGD; MGI:97848; Ragi.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR001841; Znf_ring.
Pfam; PF00097; zf-C3HC4; 1.
SMART; SM00184; RING; 1.
SMART; SM00355; Znf_C2H2; 2.
PROSITE; PS00518; ZF_RING_1; 1.
PROSITE; PS00899; ZF_RING_2; 1.
KW      Hydrolase; Endonuclease; Nuclear protein; DNA-binding;
KW      DNA recombination; Zinc-finger; Metal-binding; 3D-structure.
FT      ZN_FING 290 329
FT      ZN_RING 329 444
FT      DNA_BIND 389 444
FT      METAL 266 266
FT      METAL 270 270
FT      METAL 290 290
FT      METAL 293 293
FT      METAL 295 295
FT      METAL 305 305
FT      METAL 307 307
FT      METAL 310 310
FT      METAL 313 313
FT      METAL 325 325
FT      METAL 328 328
FT      METAL 355 355
FT      METAL 360 360
FT      METAL 372 372
FT      METAL 376 376
FT      METAL 600 600
FT      METAL 708 708
FT      ACT_SITE 962 962

```

```

FT MUTAGEN 546 546 D->A: LOSS OF DNA-BINDING.
FT MUTAGEN 560 560 D->Q: LOSS OF DNA-BINDING.
FT MUTAGEN 597 597 E->Q: IMPAIRED CLEAVAGE.
FT MUTAGEN 600 600 D->N: LOSS OF CLEAVAGE (BOTH NICKING AND
FT HAIRPIN FORMATION).
FT MUTAGEN 600 600 D->A: LOSS OF CLEAVAGE AND STRAND
FT TRANSFER ACTIVITIES
FT MUTAGEN 708 708 D->N: LOSS OF CLEAVAGE (BOTH NICKING AND
FT HAIRPIN FORMATION).
FT MUTAGEN 708 708 D->A: LOSS OF CLEAVAGE AND STRAND
FT TRANSFER ACTIVITIES.
FT MUTAGEN 709 709 E->Q: IMPAIRED CLEAVAGE.
FT MUTAGEN 709 709 E->A: IMPAIRED CLEAVAGE (DEFECTIVE IN
FT HAIRPIN FORMATION).
FT MUTAGEN 713 713 E->A,C: IMPAIRED CLEAVAGE (BOTH NICKING
FT AND HAIRPIN FORMATION).
FT MUTAGEN 719 719 E->Q: IMPAIRED CLEAVAGE.
FT MUTAGEN 792 792 D->N: IMPAIRED CLEAVAGE.
FT MUTAGEN 811 811 E->Q: LOSS OF DNA-BINDING.
FT MUTAGEN 811 811 E->A: IMPAIRED CLEAVAGE.
FT MUTAGEN 959 959 E->Q: IMPAIRED CLEAVAGE.
FT MUTAGEN 962 962 E->Q: LOSS OF CLEAVAGE (BOTH NICKING AND
FT HAIRPIN FORMATION).
FT MUTAGEN 986 986
FT HELIX 266 268
FT TURN 269 269
FT HELIX 273 275
FT STRAND 276 276
FT HELIX 282 287
FT TURN 288 288
FT STRAND 289 289
FT TURN 291 293
FT STRAND 296 296
FT STRAND 300 302
FT TURN 304 305
FT STRAND 308 310
FT HELIX 311 320
FT STRAND 321 321
FT STRAND 324 324
FT TURN 326 328
FT STRAND 331 331
FT HELIX 334 336
FT STRAND 338 338
FT STRAND 342 350
FT STRAND 352 354
FT TURN 358 359
FT STRAND 363 365
FT HELIX 366 374
FT TURN 375 375
SQ SEQUENCE 1040 AA; 119160 MW; 034317A931A0D547 CRC64;

Alignment Scores:
Pred. No.: 4.76 Length: 1040
Score: 70.50 Matches: 23
Percent Similarity: 45.95% Conservative: 11
Best Local Similarity: 21.08% Mismatches: 21
Query Match: 10.04% Indels: 19
DB: 1 Gaps: 4

M64347 (1-385) x RAG1_MOUSE (1-1040)
QY 84 CCGCCCTGTCGAGGTCGAGTTATTAGATTACAGTTTATATATCTATATA 143
DB 464 ProlaValCysLeu-----AlaIleArgValAsnThrPheLeu----- 476
QY 144 TATAATTATTAGTTTACAGATGAT----- 173
DB 477 ---SerCysSerGlnTyrHisLysMetTyrArgThrValLysAlaIleThrGlyArgGln 495
QY 174 TTGTTGTAGACTTAACACTTCTTACGAATGCTTCTAGAGTTTATAGCTGGA 233
DB 496 IlePheGlnProLeuHisAlaLeuArgAsnAlaGluLysValLeuLeuProGlyTyrHis 515

```

```

QY 234 CCTTCAAGCTTGAGGAGCCGTCGATTCAGTTGGTTCG 275
DB 516 ProPheGlu---TrpGlnProLeuLysAsnValSer 528

RESULT 4
RAG1_RABIT
ID RAG1_RABIT STANDARD; PRT; 1042 AA.
AC P34088;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE V(D)J recombination activating protein 1 (RAG-1).
GN RAG1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93354283; PubMed=8350872;
RA Fuschiotti P., Harindranath N., Mage R.G., McCormack W.T.,
RA Dhanarajan P., Roux X.H.;
RT "Recombination activating genes-1 and -2 of the rabbit: cloning and
RT characterization of germline and expressed genes.";
RL Mol. Immunol. 30:1021-1032(1993).
CC -!- FUNCTION: During lymphocyte development, the genes encoding
CC immunoglobulins and T cell receptors are assembled from variable
CC (V), diversity (D), and joining (J) gene segments. This
CC combinatorial process, known as V(D)J recombination, allows the
CC generation of an enormous range of binding specificities from a
CC limited amount of genetic information. The RAG1/RAG2 complex
CC initiates this process by binding to the conserved recombination
CC signal sequences (RSS) and introducing a double-strand break
CC between the RSS and the adjacent coding segment. These breaks are
CC generated in two steps, nicking of one strand (hydrolysis),
CC followed by hairpin formation (transesterification). RAG1/2 has
CC also been shown to function as a transposase in vitro, and to
CC possess RSS-independent endonuclease activity (end processing) and
CC hairpin opening. RAG1 alone can bind to RSS but stable, efficient
CC binding requires RAG2. All known catalytic activities require the
CC presence of both proteins (By similarity).
CC -!- COFACTOR: Binds 1 magnesium or manganese ion per subunit (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Thymus.
CC -!- DOMAIN: The specific binding to the nonamer RSS motif is mediated
CC by the nonamer binding domain (NBD) (By similarity).
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -!- SIMILARITY: Contains 1 nonamer binding (NBD) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M77666; AAA03025.1; -.
CC PIR; S42511; S42511.
CC HSP; P15919; 1RMD.
CC InterPro; IPR001841; Znf_ring.
CC Pfam; PF00097; zf-C3HC4_1.
CC SMART; SM00184; RING_1.1.
CC PROSITE; PS00518; ZF_RING_1; 1.
CC PROSITE; PS00089; ZF_RING_2; 1.
CC Hydrolase; Endonuclease; Nuclear protein; DNA-binding;
CC DNA recombination; Zinc-finger; Metal-binding.
CC ZN FING 292 331 RING-TYPE.
CC FT DN5 BIND 391 446 NBD.
CC FT METAL 602 602 DIVALENT METAL ION (BY SIMILARITY).
CC FT METAL 710 710 DIVALENT METAL ION (BY SIMILARITY).
CC FT ACT_SITE 964 964 BY SIMILARITY.

```

```

SQ SEQUENCE 1042 AA; 119006 MW; 136C6286C6E22FAE CRC64;

Alignment Scores:
Pred. No.: 4.76 Length: 1042
Score: 70.50 Matches: 23
Percent Similarity: 45.95% Conservative: 11
Best Local Similarity: 31.08% Mismatches: 21
Query Match: 10.04% Indels: 19
DB: 1 Gaps: 4

M64347 (1-385) x RAG1_RABIT (1-1042)
QY 84 CCGGCGCTGTGTCAGGTCCTCGATGTTATTAGATGTTTACAAAGTTTATATATATATATATATA 143
Db 466 ProAlaValCysLeu-----AlaIleArgValAsnThrPheLeu----- 478
QY 144 TATAATTTATGAGTTTTCACAGATGAT- 173
Db 479 ---SerCysSerGlnTyrHisLysMetTyrArgThrValLysAlaIleThrGlyArgGln 497
QY 174 TTGTTGAGACTTAAACATCTTACCAATGCTTCTAGAGTTTATAGCTGACTGCTA 233
Db 498 IlePheGlnProIleHisAlaLeuA-gAsnAlaGluLysValLeuLeuProGlyTyrHis 517
QY 234 CTTTCAAAGCTTGGAGGGAAGCGGTGAATTCAGTTGGTTTCG 275
Db 518 ProPheGlu---TrpGlnProProLeuLysAsnValSerSer 530

RESULT 5
LNK_MOUSE
ID LNK_MOUSE STANDARD; PRT; 548 AA.
AC O09039;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lymphocyte specific adaptor protein Lnk (Signal transduction protein
LNK) (Lymphocyte adaptor protein).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=FVB/N; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Joquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length

```

```

RT human and mouse cDNA sequences.":
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).
CC -!- FUNCTION: Links T-cell receptor activation signal to phospholipase
CC C-gamma-1, GEB-2 and phosphatidylinositol 3-kinase (By
CC similarity).
CC -!- PTM: Tyrosine phosphorylated (By similarity).
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 SH2 domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U89992; AAB58580.2; -.
CC EMBL; U89993; AAB58581.1; -.
CC EMBL; BC006759; AAB06759.1; -.
CC HSSP; P29353; 1MIL.
CC MGD; MGI:93559; Lnk.
CC InterPro; IPR001849; PH.
CC InterPro; IPR000980; SH2.
CC Pfam; PF00017; SH2; 1.
CC PRINTS; PR00401; SH2DOMAIN.
CC ProDom; PD000093; SH2; 1.
CC SMART; SM00233; PH; 1.
CC SMART; SM00252; SH2; 1.
CC PROSITE; PS50001; SH2; 1.
CC Phosphorylation; SH2 domain. PH.
FT DOMAIN 168 279
FT DOMAIN 336 434 SH2.
SQ SEQUENCE 548 AA; 60487 MW; D21DCB46185962B8 CRC64;

Alignment Scores:
Pred. No.: 6.7 Length: 548
Score: 69.00 Matches: 40
Percent Similarity: 35.80% Conservative: 18
Best Local Similarity: 24.69% Mismatches: 46
Query Match: 10.00% Indels: 58
DB: 1 Gaps: 10

M64347 (1-385) x LNK_MOUSE (1-548)
QY 380 CTGGCCCATGCTACTGGGCCCAAGAGACACCTGAGCCATGGCCC----- 330
Db 39 LeuAlaArgGlnTyrTrpLeuPheAlaArgAlaHisProGlnPro---ProArgAlaAsp 57
QY 329 -----TGACGCAACCAAGGAC 312
Db 58 LeuValSerLeuGlnPheAlaGluLeuPheGlnArgHisPheCysArgGluValArgGlu 77
QY 311 AGCTGCCAGACTCAGGCCCAAGTAAACAGTACAGAACCAACCACTGAATTCACGGCTTC 252
Db 78 SerLeu-----AlaGlyProProGlyHisAspTyrArgAlaThr-----Ala 91
QY 251 CTTCAAGCTTTGAAGAGTAGGATGAGTCAGGCTATATAAACTCTAGAGCATTCGCTAAGAA 192
Db 92 ProPro-----ArgProAlaLeuProLysAlaArgSerSer-----Glu 104
QY 191 GGTGTA-----AGTCTACCAAAATACATCTTGTATAAAACTCAATAA 150
Db 105 AspLeuGlyProArgProAlaCysAlaLeuGlnHisLeuArgArgGlyLeuArgGln--- 123
QY 149 ATTATATATATAGATATATATAA-----CTTGTGTA 120
Db 124 -----LeuPheArgArgSerAlaGlyGluLeuProGlyAlaThrSer 138
QY 119 ACATCTTAATACATCGGAACCTGCACACAGGCCCGCCCTCCCTGGAAACCGTCTCCCT 60
Db 139 AspThrAsnAspIleAspThrThrAlaAlaSerArgProGlyProAlaArgLysLeuLeu 158

```

QY 59 GCCTGG-----GACACACAGCAATTAAGAAGATTGTATGAATACACAG 15
 Db 159 ProTPGlyLeuArgGluProThrGluAlaLeuGlyGluValLeuArgTySer 178
 QY 14 CTGTGT 9
 Db 179 LeuAla 180

RESULT 6
 IRX5_MOUSE
 ID IRX5_MOUSE STANDARD; PRT; 484 AA.
 AC Q9JUK4; Q9JUL5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Iroquois-class homeodomain protein IRX-5 (Iroquois homeobox protein
 DE 5).
 GN IRX5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20284900; PubMed=10822268;
 RA Bosse A., Stoykova A., Nieselt-Struwe K., Chowdhury K., Copeland N.G.,
 RA Jenkins N.A., Gruss P.;
 RT "Identification of a novel mouse Iroquois homeobox gene, Irx5, and
 RT chromosomal localisation of all members of the mouse Iroquois gene
 RT family";
 RL Dev. Dyn. 218:160-174 (2000).
 RN [2]

SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Lung;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli P., Suzuki R., Tonita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., Fujita M., Gariboldi M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Kamiya M., Lee N.H.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Yushaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 403:685-690 (2001).
 RN [3]

SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra W.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [4]

SEQUENCE OF 1-442 FROM N.A.
 RX MEDLINE=20171054; PubMed=10704856;
 RA Cohen D.R., Cheng C.W., Cheng S.H., Hui C.C.;
 RT "Expression of two novel mouse Iroquois-class homeobox genes during
 RT neurogenesis";
 RL Mech. Dev. 91:317-321 (2000).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- SIMILARITY: Belongs to the TALE/IRO homeobox family.
 CC -1- SIMILARITY: Contains 1 homeobox domain.
 CC -1- CAUTION: Ref.4 sequence differs from that shown due to a
 CC frameshift in position 430.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF230074; AAF42871.1; -
 DR EMBL; AK004747; BAB23528.1; -
 DR EMBL; BC051959; AAH51959.2; -
 DR EMBL; BC056994; AAH56994.1; -
 DR EMBL; AF165985; AAF63955.1; ALT_FRAME.
 DR HSSP; P41778; 1DU6.
 DR TRANSFAC; T04283; -
 DR MGD; MGI:1859086; Irx5.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR003893; Iroquois_homeo.
 DR Pfam; PF00046; homeobox; 1.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR SMART; SM00548; IRO; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 DR Homeobox; DNA-binding; Nuclear protein.
 KW DNA_BIND 112 174 HOMEBOX (TALE-TYPE).
 FT DOMAIN 72 78 POLY-ALA.
 FT DOMAIN 185 191 POLY-GLU.
 FT DOMAIN 319 327 POLY-PRO.
 FT DOMAIN 351 354 POLY-GLY.
 SQ SEQUENCE 484 AA; 50755 MW; 4C9AB3ED9992EDC9 CRC64;

Alignment Scores:
 Pred. No.: 9.93 Length: 484
 Score: 67.50 Matches: 31
 Percent Similarity: 31.93% Conservative: 7
 Best Local Similarity: 26.05% Mismatches: 36
 Query Match: 9.78% Indels: 45
 DB: 1 Gaps: 6

M64347 (1-385) x IRX5_MOUSE (1-484)

QY 376 CCACCATGCTGGGCCCCAAGAGAGACACACCTTG-----AGCCATGGC 332
 Db 361 ProProCysProGlyProMetGlyGlyGlnThrLeuGlyGlySerArgAlaSerProAla 380
 QY 331 CTGTGAGGACAGGAGGACAGCTGCCAGACTAGGCCCCA---GTAACAGTACAGAC 275
 Db 381 ProAlaProAlaArgSerProSerAlaGlnCysProPheProGlyGlyThrValLeuSer 400
 QY 274 GAACCAACTGAATTCACCGCTTCCTCCCAAGCTTTGAAGGTAGCAGTCCAGGCTATAAA 215
 Db 401 ArgProLeuTyThrAlaProPheTy-----ProGlyTyThr 414

```

DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nonstructural protein NS3.
GN NS-3.
OS Rice stripe virus (isolate T) (RSV).
OC Viruses; ssRNA negative-strand viruses; Tenuivirus.
NCBI_TaxID=36394;
[1]_TaxID=36394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91202110; PubMed=2016591;
RA Zhu Y., Hayakawa T., Toriyama S., Takahashi M.;
RT "Complete nucleotide sequence of RNA 3 of rice stripe virus: an
RT ambisense coding strategy.";
RL J. Gen. Virol. 72:763-767(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X53563; CAA37634.1; -
CC InterPro; IPR007974; Tenuivirus_NS3.
CC Pfam; PF05310; Tenuivirus_NS3; 1.
CC Nonstructural protein.
KW Nonstructural protein.
SQ SEQUENCE 211_AA; 23874 MW; 8C06ACA56E04A20 CRC64;
-----
Alignment Scores:
Pred. No.: 10.4 Length: 211
Score: 67.00 Matches: 14
Percent Similarity: 46.34% Conservative: 5
Best Local Similarity: 34.15% Mismatches: 14
Query Match: 9.71% Indels: 8
DB: 1 Gaps: 1
M64347 (1-385) x VN53_RSVT (1-211)
QY 359 CCAAGAAGAGACACACCTGAGCCATGCCCTGAGGCAAGCAAGGACAGCTGCCACGAC 300
Db 113 ProArgProProSerThrAsnCysTyrThrCysArgMetSerLysAspAsnLeuProPhe 132
QY 299 TCAGGGCCCGAGTAACTACAGTACAGACGACGACCACTGAATTCACGGCTTCCTCCAGCTTT 240
Db 133 ThrValProSer-----VallysGlyPheProProAspAla 144
QY 239 GAA 237
Db 145 Glu 145
-----
RESULT 9
CD68_MOUSE
ID _CD68_MOUSE STANDARD; PRT; 326 AA.
AC P31956;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Macrosialin precursor (Cd68 antigen).
GN CD68.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OC NCEI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BALB/c; TISSUE=Macrophage;
RA MEDLINE=93252841; PubMed=8486654;
RA Holmes C.L., da Silva R.P., Fawcett J., Gordon S., Simmons D.L.;
RT "Macrosialin, a mouse macrophage-restricted glycoprotein, is a member
RT of the lamp/19p family.";
RL J. Biol. Chem. 268:9661-9666(1993).
RN [2]

```

RP SEQUENCE FROM N.A.
RX MEDLINE=98148094; PubMed=9479000;
RA Li A.C., Guidez F.R.B., Collier J.G., Glass C.K.;
RT "The macroalgalin promoter directs high levels of transcriptional
activity in macrophages dependent on combinatorial interactions
between PU.1 and c-Jun.";
RL J. Biol. Chem. 273:5389-5399(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=98317533; PubMed=9653646;
RA Jiang Z., Shih D.M., Xia Y.R., Luis A.J., de Beer F.C.,
RA de Villiers W.J.S., van der Westhuyzen D.R., de Beer M.C.;
RT "Structure, organization, and chromosomal mapping of the gene encoding
macrostalinalin, a macrophage-restricted protein.";
RL Genomics 50:199-205(1998).
CC -!- FUNCTION: Could play a role in phagocytic activities of tissue
macrophages, both in intracellular lysosomal metabolism and
extracellular cell-cell and cell-pathogen interactions. Bind to
tissue- and organ-specific lectins or selectins, allowing homing
of macrophage subsets to particular sites. Rapid recirculation of
CD68 from endosomes, lysosomes to the plasma membrane may allow
macrophages to crawl over selectin bearing substrates or other
cells.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ENDOSOMAL OR
LYSOSOMAL (LONG VARIANT) AND TO A LESSER EXTENT ON THE CELL
SURFACE (SHORT VARIANT).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P31996-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P31996-2; Sequence=VSP 003043;
CC -!- TISSUE SPECIFICITY: Expressed in tissue macrophages and to a
lesser extent in dendritic cells.
CC -!- PTM: N- and O-glycosylated.
CC -!- SIMILARITY: Belongs to the LAMP family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).
CC
CC EMBL; X68273; CAA48334.1; -;
DR EMBL; AF039399; AAC40056.1; -;
DR EMBL; AF022651; AAC40151.1; -;
DR FIR; A46676; A46676.
DR MGD; MGI:88342; Cd68.
DR Pfam; PF01299; LAMP; 1.
DR InterPro; IPR002000; LAMP.
DR PRINTS; PR00336; LYASASOCTDMP.
DR PROSITE; PS00311; LAMP 2; 1.
KW Transmembrane; Glycoprotein; Signal; Lysosome; Repeat; Antigen;
KW Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 326
FT DOMAIN 21 291
FT DOMAIN 292 316
FT TRANSMEM 317 326
FT DOMAIN 317 326
FT DOMAIN 21 109
FT DOMAIN 110 121
FT REPEAT 44 49
FT REPEAT 50 64
FT REPEAT 65 72
FT REPEAT 73 88
FT REPEAT 139 177
FT DISULFID 249 286
FT DISULFID 60 60
FT CARBOHYD 84 84
FT CARBOHYD 97 97

FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 319 326 Missing (in isoform Short).
SQ SEQUENCE 326 AA; 34818 MW; AB7203A9A7EA47BA CRC64;
/FTID=VSP_003043.
Alignment Scores:
Pred. No.: 10.9 Length: 326
Score: 67.00 Matches: 32
Percent Similarity: 35.85% Conservative: 6
Best Local Similarity: 30.19% Mismatches: 33
Query Match: 9.71% Indels: 35
DB: 1 Gaps: 6
M64347 (1-385) x CD68_MOUSE (1-326)
QY 367 ACTGGGCCCCAAGAGAGACACCCCTGAGCCATGGCCCTGCAGGCAAGCAAGGACGCT 308
Db 48 ThrThrSerHisArgProThrThrThrSerHisGlyAsnVal-----ThrVal 63
QY 307 GCCCAGACTCAGGGCCCGAGTAACAGTACAGAACCAACCACTGAATTCACGGCTTCCCTC 248
Db 64 HisThrSerSerGlyProThrThrValThrHisAsnProAlaThrThr----- 80
QY 247 CAAGCTTTGAAAGGTAGCAGTCCAGCTTAAACACTCTAGAAGCATTCGCTAAGAGTGT 188
Db 81 -----SerHisGlyAsnAlaThrIle-SerHisAla-----ThrVa 92
QY 187 TAAGTCTACAACAAT-----ACATCTGTAAAAAACTCAATAAATATATATATAG 137
Db 92 lSerProThrThrAsnGlyThrAlaThrSerProArgSerThr----- 107
QY 136 ATATATATAAATTTGTAACATCTAATAACATCGGACCTGCACAGCGCGGCCCTCC 77
Db 108 -----ValGlyProHisProGlyProProProPr 117
QY 76 CTGGAACCGCTCTCCC 61
Db 117 O---SerProSerPro 121
RESULT 10
DVR2 HUMAN STANDARD; PRT; 528 AA.
AC Q92630;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Dual-specificity tyrosine-phosphorylation regulated kinase 2
DE (EC 2.7.1.-).
GN DYRK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98421512; PubMed=9748265;
RA Becker W., Weber Y., Wetzel K., Eirnbter K., Tejedor F.J.,
RA Joost H.-G.;
RT "Sequence characteristics, subcellular localization, and substrate
specificity of DYRK-related kinases, a novel family of dual
specificity protein kinases.";
RL J. Biol. Chem. 273:25893-25902(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

M64347 (1-385) x DYR2_HUMAN (1-528)

QY 117 TGTTCACAAAGTTTATATATATCTTATATATATAAATTATTGAGTGGTTTTTACAAGATGTATTGTG 176

DB 300 CysTy rGluHisGlnArgValTy rThrTy rIleGlnSerArgPheTy r-ArgAlaProgl 319

QY 177 TTGTAGACTTAACACTTCTTACGCAGATGCTTCTAGAGTGTATTAGCTGGACTGCTACCT 236

DB 319 uValllleuGlyalaArgTy rGlyMetProfileAspMetTrpSerLeuGlyCysileLe 339

QY 237 TTCAAAGCTTGCGAGGAAGCCGTGAATTCAGTTGGTTCGTTCTGTACTGTACTGGGCC 296

DB 339 alaGluleu-----LeuThrGlyTy rProleuLeuProGlyCl 352

QY 297 TGAGTCTGGGAGCTGTCCCTTGGTTCCCTG 327

DB 352 uAspgluglyAsp---GlnLeuAlaCysMet 361

RESULT 11

ATSS_HUMAN STANDARD; PRF; 930 AA.

ID ATSS_HUMAN Q9UNA0; Q9UKP2; AC DT 16-OCT-2001 (Rel. 40, Created) DT 16-OCT-2001 (Rel. 40, Last sequence update) DT 10-OCT-2003 (Rel. 42, Last annotation update) DE ADAMTS-5 precursor [EC 3.4.24.-] (A disintegrin and metalloproteinase with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TSS) (Aggrecanase-2) DE (ADMP-2) (ADAM-TS 11). DE ADAMTS5 OR ADMP2 OR ADAMTS11. GN Homo sapiens (Human). OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. OX NCBI_TaxID=9606; IL TaxID=9606; RN SEQUENCE FROM N.A. RC TISSUE=Liver; RX MEDLINE=99367476; PubMed=10438522; RA Abbaszade I., Liu R.-Q., Yang F., Rosenfeld S.A., Ross O.H., Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M., Wynn R., Duke J.L., George H.J., Hillman M.C. Jr., Murphy K., Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R.; Nagase H., Ito Y., Newton R.C., Magolda R.L., Trzakos J.M., Hollis G.F., Arner E.C., Burn T.C.; RA "Cloning and characterization of ADAMTS11, an aggrecanase from the ADAMTS family." RT J. Biol. Chem. 274:23443-23450(1999). {2} SEQUENCE FROM N.A. RN MEDLINE=20289799; PubMed=10830953; RX Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y., Soeda E., Ohki M., Takagi T., Sakaki Y., Taubien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schillhaber M., Schudy A., Zimmermann W., Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S., Shintani A., Sasaki T., Nagamine K., Matsuyama S., Antonarakis S.E., Minoshima S., Shimizu N., Nordstok G., Hornischer K., Brandt P., Schaefer M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H., Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E., Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., Yaspo M.-L.; RA "The DNA sequence of human chromosome 21." RT Nature 405:311-319(2000). {3} SEQUENCE OF 413-930 FROM N.A. RN RN TISSUE=Fetal brain; RC MEDLINE=99395124; PubMed=10464288; RX Hurskainen T.I., Hirohata S., Seldin M.F., Apte S.S.; RA "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of zinc metalloproteases." RT J. Biol. Chem. 274:25555-25563(1999). CC -!- FUNCTION: Cleaves aggrecan, a cartilage proteoglycan, and may be

RP SEQUENCE OF 413-530 FROM N.A.
 RZ TISSUE:Fetal brain;
 RX MEDLINE:9395124; PubMed:10464288;
 RA Furukainen T.L., Hirschata S., Seidin M.F., Apté S.S.;
 RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of
 RI zinc metalloproteases.";
 RJ J. Biol. Chem. 274:25555-25563 (1999).
 RL -!- FUNCTION: Cleaves aggrecan, a cartilage proteoglycan, and may b

involved in its turnover. May play an important role in the destruction of aggrecan in arthritic diseases. May play a role in proteolytic processing mostly during the peri-implantation period.

CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-Ala-393 site.

CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).

CC -!- TISSUE SPECIFICITY: Expressed at low level, primarily in placenta but also in other tissues, such as heart and brain, and also cervix, uterus, bladder, esophagus, rib cartilage, chondroblastoma, fibrous tissue and joint capsule from an arthritic patient.

CC -!- DOMAIN: The spacer domain and the TSP type-1 domains are important for a tight interaction with the extracellular matrix.

CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By similarity).

CC -!- SIMILARITY: Belongs to peptidase family M12B.

CC -!- SIMILARITY: Contains 1 disintegrin-like domain.

CC -!- SIMILARITY: Contains 2 TSP type-1 domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

DR EMBL; AF142099; AAD49577.1; -

DR EMBL; AP001698; BAA95504.1; -

DR EMBL; AP001697; BAA95503.1; -

DR EMBL; AF141293; AAF02493.1; -

DR HSP; QPWS35; IBUD.

DR MEROPS; M12.225; -

DR Genew; HGNC:221; ADAMTSS.

DR MIM; 605007; -

DR GO; GO:0005578; C:extracellular matrix; TAS.

DR GO; GO:0005178; F:integrin binding; TAS.

DR GO; GO:0008237; F:metalloproteinase activity; TAS.

DR GO; GO:0006508; F:proteolysis and peptidolysis; TAS.

DR InterPro; IPR001762; Disintegrin.

DR InterPro; IPR006025; Pept_M_Zn_BS.

DR InterPro; IPR001590; Peptidase M12B.

DR InterPro; IPR002870; Peptidase_M12B_N.

DR InterPro; IPR000884; TSP1.

DR InterPro; IPR008085; TSP1.

DR Pfam; PF01562; Pep_M12B_propep; 1.

DR Pfam; PF01421; Reprolysin; 1.

DR Pfam; PF00090; tsp_1; 2.

DR PRINTS; PR01705; TSP1REPEAT.

DR SMART; SM00209; TSP1; 2.

DR PROSITE; PS00215; ADAM_MEPRO; 1.

DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.

DR PROSITE; PS0214; DISINTEGRIN_2; FALSE_NEG.

DR PROSITE; PS50092; TSP1; 2.

DR PROSITE; PS00142; ZINC_PROTEASE; 1.

KW Hydrolase; Metalloproteinase; Zinc; Signal; Glycoprotein; Zymogen; Repeat; Extracellular matrix.

KW SIGNAL; 1 16 POTENTIAL.

FT PROPEP 17 261 POTENTIAL.

FT CHAIN 262 930 ADAMTS-5.

FT DOMAIN 262 484 METALLOPROTEINASE.

FT DOMAIN 485 566 DISINTEGRIN-LIKE.

FT DOMAIN 567 622 TSP TYPE-1 1.

FT DOMAIN 624 731 CYS-RICH.

FT DOMAIN 732 874 SPACER.

FT DOMAIN 875 929 TSP TYPE-1 2.

FT DOMAIN 37 41 POLY-ALA.

FT DOMAIN 257 261 POLY-ARG.

FT SITE 209 209 CYSTEINE SWITCH (POTENTIAL).

FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT_SITE 411 411 BY SIMILARITY.

FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).

FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 728 728 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 807 807 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CONFLICT 138 138 A -> G (IN REF. 2).

FT CONFLICT 614 614 R -> H (IN REF. 3).

FT CONFLICT 692 692 P -> L (IN REF. 2).

SQ SEQUENCE 930 AA; 101715 MW; B54281502F28193B CRC64;

Alignment Scores:

Pred. No.:	Length:	930
Score:	12.2	Matches: 38
Percent Similarity:	30.22%	Conservative: 17
Best Local Similarity:	20.88%	Mismatches: 54
Query Match:	9.71%	Indels: 73
DB:	1	Gaps: 7

M64347 (1-385) x ATSS_HUMAN (1-930)

QY 379 TGCCACCATGCCT----- 365

Db 452 TpsrLysCysThrSerAlaThrIleThrGluPheLeuAspGlyHisGlyAsnCys 471

QY 364 -----GGGCCCCAAGAA----- 353

Db 472 LeuLeuAspLeuProArgLysGlnIleLeuGlyProGluGluLeuProGlyGlnThrTyr 491

QY 352 -----GAGACACCCCTGAGCCATGCCCT-----GCAGGCAAGCAA 317

Db 492 AspAlaThrGlnGlnCysAsnLeuThrPheGlyProGluTyrSerValCysProGlyMet 511

QY 316 GGGCAGCTGCCCAGACTCAGGGCCCTACAGTACAGACGACGACCACTGAATTCACG 257

Db 512 AspValCysAlaArgLeuTyrCysAlaValValArgGlnGlyGlnMetValCysLeuThr 531

QY 256 GTTCTCCCTCAAGCTTGAAGAGTAGCAGCTCAGCTAT----- 218

Db 532 LysLysLeuProAlaValGlnGlyThrProCysGlyLysGlyArgIleCysLeuGlnGly 551

QY 217 AAACTCTAGAAGCATTCGTTAAGTGTAACTTACACA----- 175

Db 552 LysCysValAspLysThrLysLysLys-TyrTyrSerThrSerSerHisGlyAsnTrpG1 571

QY 174 -----AATACATCTTGTAAATAAATACTCAATAAATATATATA 140

Db 571 YserTrpGlySerTrpGlyGlnCysSerArgSerCysGlyGlyGlyValGlnPheAlaTy 591

QY 139 TAGATATATATAAATTGTAACATTAATAACATCGGAACCTGCACAGGGCCGGCCCC 80

Db 591 rargHis-----Cysasn-----AsnProAlaP 599

QY 79 TCCTCGGAACCGTCTCCCTGCTCGGACACACAGCAATTAAGAAGATTGTATCAAAAT 20

Db 599 oArganLanGlyArgTyrCysThrGlyLysArgAlaIleTyrArgSerCysSerLeuMe 619

QY 19 ACCA 16

Db 619 tPro 620

RESULT 12

TCOF_HUMAN

ID TCOF_HUMAN STANDARD; PRT; 1411 AA.

AC Q13428; Q99408; Q99860;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Treacle protein (Treacher Collins syndrome protein).

DE TCOF1.

GN Homo sapiens (Human).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

EMBL; U84655; AAC51185.1; JOINED.
EMBL; U84656; AAC51185.1; JOINED.
EMBL; U84657; AAC51185.1; JOINED.
EMBL; U84658; AAC51185.1; JOINED.
EMBL; U84659; AAC51185.1; JOINED.
EMBL; U84660; AAC51185.1; JOINED.
EMBL; U84661; AAC51185.1; JOINED.
EMBL; U84662; AAC51185.1; JOINED.
EMBL; U84663; AAC51185.1; JOINED.
EMBL; U79659; AAB40722.1; -.
EMBL; U79645; AAB40722.1; JOINED.
EMBL; U79646; AAB40722.1; JOINED.
EMBL; U79647; AAB40722.1; JOINED.
EMBL; U79648; AAB40722.1; JOINED.
EMBL; U79649; AAB40722.1; JOINED.
EMBL; U79650; AAB40722.1; JOINED.
EMBL; U79651; AAB40722.1; JOINED.
EMBL; U79652; AAB40722.1; JOINED.
EMBL; U79653; AAB40722.1; JOINED.
EMBL; U79654; AAB40722.1; JOINED.
EMBL; U79655; AAB40722.1; JOINED.
EMBL; U79656; AAB40722.1; JOINED.
EMBL; U79657; AAB40722.1; JOINED.
EMBL; U79658; AAB40722.1; JOINED.
Genew; HGNC:11654; TCOF1.
MIM; 606847; -.
MIM; 154500; -.
GO; GO:0005730; C:nucleolus; TAS.
GO; GO:0005215; F:transporter activity; TAS.
GO; GO:0001501; P:skeletal development; TAS.
InterPro; IPR006594; Lish.
InterPro; IPR003993; treacle.
Pfam; PF03546; treacle; 3.
PRINTS; PR01503; TREACLE.
SMART; SM00567; Lish; 1.
PROSITE; PS00896; LISH; 1.
Disease mutation; Polymorphism.
DOMAIN 6 38
DOMAIN 89 97 POLY-GLU.
DOMAIN 204 207 POLY-SER.
DOMAIN 616 619 POLY-SER.
DOMAIN 919 924 POLY-SER.
DOMAIN 1285 1289 POLY-LYS.
DOMAIN 1375 1386 POLY-LYS.
DOMAIN 1398 1405 POLY-LYS.
VARIANT 53 53 W -> R (in TCS).
VARIANT 439 439 P -> L.
VARIANT 810 810 A -> V.
VARIANT 1313 1313 A -> V (in dbSNP:15251).
VARIANT 1355 1355 P -> G.
VARIANT 1312 1312 K -> Q (in REF. 2).
CONFLICT 1411 AA; 144312 MW; 3880203D985C2699 CRC64;
SEQUENCE

nment Scores:
. No.: 12.7 Length: 1411
e: 67.00 Matches: 23
ent Similarity: 43.08% Conservative: 5
Local Similarity: 35.38% Mismatches: 27
y Match: 9.71% Indels: 10
1 Gaps: 4

47 (1-385) x TCOF_HUMAN (1-1411)
376 CCACCATGCTGGGCCCAAGAGAGACACCTGAGCATCGGCC
752 PProCvValArgAsnProGlnAsnSerThrValLeuAlaArgGlyP
325 -----GGCAAGCAA---GGGACAGCTGCCAGACTCAG---GGCCO

```


FT	DISULFID	226	BY SIMILARITY.	242
FT	DISULFID	244	BY SIMILARITY.	253
FT	DISULFID	260	BY SIMILARITY.	271
FT	DISULFID	265	BY SIMILARITY.	280
FT	DISULFID	282	BY SIMILARITY.	291
FT	DISULFID	298	BY SIMILARITY.	311
FT	DISULFID	305	BY SIMILARITY.	320
FT	DISULFID	322	BY SIMILARITY.	331
FT	DISULFID	338	BY SIMILARITY.	349
FT	DISULFID	343	BY SIMILARITY.	358
FT	DISULFID	360	BY SIMILARITY.	369
FT	DISULFID	375	BY SIMILARITY.	386
FT	DISULFID	380	BY SIMILARITY.	397
FT	DISULFID	399	BY SIMILARITY.	408
FT	DISULFID	415	BY SIMILARITY.	428
FT	DISULFID	422	BY SIMILARITY.	437
FT	DISULFID	439	BY SIMILARITY.	448
FT	DISULFID	455	BY SIMILARITY.	466
FT	DISULFID	460	BY SIMILARITY.	475
FT	DISULFID	477	BY SIMILARITY.	486
FT	DISULFID	493	BY SIMILARITY.	504
FT	DISULFID	498	BY SIMILARITY.	513
FT	DISULFID	515	BY SIMILARITY.	524
FT	DISULFID	531	BY SIMILARITY.	542
FT	DISULFID	536	BY SIMILARITY.	551
FT	DISULFID	553	BY SIMILARITY.	562
FT	DISULFID	569	BY SIMILARITY.	579
FT	DISULFID	574	BY SIMILARITY.	588
FT	DISULFID	590	BY SIMILARITY.	599
FT	DISULFID	606	BY SIMILARITY.	617
FT	DISULFID	611	BY SIMILARITY.	626
FT	DISULFID	628	BY SIMILARITY.	637
FT	DISULFID	644	BY SIMILARITY.	654
FT	DISULFID	649	BY SIMILARITY.	663
FT	DISULFID	665	BY SIMILARITY.	674
FT	DISULFID	681	BY SIMILARITY.	692
FT	DISULFID	686	BY SIMILARITY.	701
FT	DISULFID	703	BY SIMILARITY.	712
FT	DISULFID	719	BY SIMILARITY.	729
FT	DISULFID	724	BY SIMILARITY.	738
FT	DISULFID	740	BY SIMILARITY.	749
FT	DISULFID	756	BY SIMILARITY.	767
FT	DISULFID	761	BY SIMILARITY.	776
FT	DISULFID	778	BY SIMILARITY.	787
FT	DISULFID	794	BY SIMILARITY.	805
FT	DISULFID	799	BY SIMILARITY.	814
FT	DISULFID	816	BY SIMILARITY.	825
FT	DISULFID	832	BY SIMILARITY.	843
FT	DISULFID	837	BY SIMILARITY.	854
FT	DISULFID	856	BY SIMILARITY.	865
FT	DISULFID	872	BY SIMILARITY.	883
FT	DISULFID	877	BY SIMILARITY.	892
FT	DISULFID	894	BY SIMILARITY.	903
FT	DISULFID	910	BY SIMILARITY.	921
FT	DISULFID	915	BY SIMILARITY.	930
FT	DISULFID	932	BY SIMILARITY.	941
FT	DISULFID	986	BY SIMILARITY.	997
FT	DISULFID	991	BY SIMILARITY.	1006
FT	DISULFID	1008	BY SIMILARITY.	1017
FT	DISULFID	1024	BY SIMILARITY.	1035
FT	DISULFID	1029	BY SIMILARITY.	1044
FT	DISULFID	1046	BY SIMILARITY.	1055
FT	DISULFID	1062	BY SIMILARITY.	1073
FT	DISULFID	1067	BY SIMILARITY.	1082
FT	DISULFID	1084	BY SIMILARITY.	1093
FT	DISULFID	1100	BY SIMILARITY.	1121
FT	DISULFID	1115	BY SIMILARITY.	1130

Alignment Scores:

Pred. No.:

15.5

Score:

66.50

Percent Similarity:

39.06%

Best Local Similarity:

31.25%

Length:

2524

Matches:

20

Conservative:

5

Mismatches:

26

Query Match:	9.64%	Indels:	13
DB:	1	Gaps:	3
M64347 (1-385) x NOTC_XENLA (1-2524)			
QY	201	TCGTTAAGTGTGTTAAGTCTACAAACAATACATCTTGTAAACTCA-----ATA	151
Db	176	CysLysGlnAspIleAsnGluCysSerGlnAsnProCysLysAsnGlyGlnCysile	195
QY	150	AATTATATATATATATATATAAATCTGTAACATCTATAACATCGGAACCTGCACACA	91
Db	196	AsnGluPheGlySerTyrArgCysThrCys-----GlnAsnArgPheThr	210
QY	90	GGG-----CCGGCCCTCCCTGGAAACCGTCTCCTCGCTGGGACACACA	46
Db	211	GlyArgAsnCysAspGluProTyrValProCysAsnProSerProCysLeuAsnGlygly	230
QY	45	GCAATTAGAAGA	34
Db	231	ThrCysArgGln	234
RESULT 14			
AAAI HUMAN			
ID	AAAI HUMAN	STANDARD;	PRT; 523 AA.
AC	Q9NS82;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Asc-type amino acid transporter 1 (Asc-1).		
GN	SLC7A10 OR ASC1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RC	SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.		
RP	TISSUE=Brain;		
RX	MEDLINE=20323545; PubMed=10863037;		
RA	Nakauchi J., Matsuo H., Kim D.K., Goto A., Chairoungdua A., Cha S.H.,		
RA	Inatomi J., Shiohara Y., Yamaguchi K., Saito I., Endou H., Kanai Y.;		
RT	"Cloning and characterization of a human brain Na+-independent		
RT	transporter for small neutral amino acids that transports D-serine		
RT	with high affinity."		
RL	Neurosci. Lett. 287:231-235(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A., AND VARIANT CSNU3 ASP-112.		
RX	MEDLINE=21400982; PubMed=11509015;		
RA	Leclerc D., Wu Q., Ellis J.R., Goodyer P., Rozen R.;		
RT	"Is the SLC7A10 gene on chromosome 19 a candidate locus for		
RT	cystinuria?"		
RL	Mol. Genet. Metab. 73:333-339(2001).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Kidney;		
RA	Bassi M.T., Borsani G., Nunes V., Palacin M.;		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RX	MEDLINE=2238257; PubMed=12477932;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 22, 2004, 10:44:15 ; Search time 91.5 Seconds
(without alignments)
2655.180 Million cell updates/sec

Title: M64347

Perfect score: 702

Sequence: 1 GACTTCAAGCAAGCTGTGTA.....GTGCATGTGGCCAGAGGTG 385

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cpn2 1/USPTO spooll_p/CANELL305B/runat_22092004_112335_14506/app_query.fasta_1.583
-DB=SPREMBL_25 -OFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40 cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=CANELL305B @CGN 1 1 211 @runat_22092004_112335_14506 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriaph:*

17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
C 1	85.5	12.4	674 16 Q8U5E6	Q8U5E6 agrobacteri

RESULT 1

Q8U5E6

ID Q8U5E6 PRELIMINARY; PRT; 674 AA.

AC Q8U5E6;

DT 01-JUN-2002 (TRENBLrel. 21, Created)

DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)

DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)

DE AGR_C 1968p.

GN AGR_C 1968.

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.

OX NCBI_TaxID=176299;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21608551; PubMed=11743194;

RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,

RA Qureshi B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,

RA Hounel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,

RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,

RA Planagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,

RA Cielo C., Slater S.

RT "Genome sequence of the plant pathogen and biotechnology agent

RT Agrobacterium tumefaciens C58."

RL Science 294:2323-2328(2001).

ALIGNMENTS

C	2	83	12.0	504	4	Q96T16
C	3	80.5	11.7	440	4	Q8IV67
C	4	80	11.6	552	4	Q86W50
C	5	76.5	11.1	283	13	Q90Z54
C	6	74.5	10.6	283	5	Q17806
C	7	73	10.6	867	3	Q96IZ7
C	8	73	10.4	955	10	Q9FQ80
C	9	72.5	10.3	474	5	Q18613
C	10	71	10.1	324	17	Q97UC7
C	11	71	10.1	447	10	Q04114
C	12	71	10.3	2114	5	Q8IQ18
C	13	71	10.1	2386	4	Q7Z7M0
C	14	70.5	10.0	360	16	Q9FKJ7
C	15	70.5	10.0	735	10	Q7XNX1
C	16	70.5	10.0	1042	6	Q9BDS2
C	17	70.5	10.0	1043	6	Q867B5
C	18	70.5	10.2	1865	10	Q9C8G1
C	19	70.5	10.2	2254	10	Q38971
C	20	70.5	10.2	2254	10	Q38970
C	21	70.5	10.2	2257	10	Q9SKV1
C	22	70	10.0	323	11	Q7TRM0
C	23	70	10.0	1193	10	Q9LDM4
C	24	70	10.1	2093	5	Q9UI31
C	25	69.5	9.9	672	10	Q7X612
C	26	69.5	9.9	712	12	Q37168
C	27	69.5	10.1	878	5	Q9VSK2
C	28	69.5	10.1	2214	5	Q20219
C	29	69	10.0	420	2	Q9F6S9
C	30	69	9.8	749	11	Q80V41
C	31	69	10.0	920	5	Q95P13
C	32	69	9.8	2090	5	Q94247
C	33	69	9.8	2153	5	Q94246
C	34	68.5	9.9	94	6	Q9BEZ9
C	35	68.5	9.8	324	13	Q90XK9
C	36	68.5	9.8	324	13	Q90XK8
C	37	68.5	9.8	347	16	Q89XK8
C	38	68.5	9.9	396	13	Q802X1
C	39	68.5	9.9	605	16	Q92X67
C	40	68.5	9.9	1188	5	Q9VED3
C	41	68.5	9.8	2185	5	Q8W6J4
C	42	68	9.7	296	2	Q51556
C	43	68	9.9	420	2	Q8GHA8
C	44	68	9.7	552	12	Q9Q6Q7
C	45	68	9.7	649	3	Q60067

Q96T16 homo sapien
Q8IV67 homo sapien
Q86W50 homo sapien
Q90Z54 Gallus gall
Q17806 caenorhabdi
Q36TZ7 neurospora
Q18613 caenorhabdi
Q97UC7 sulfolobus
Q04114 perilla fru
Q8IQ18 drosophila
Q7Z7M0 homo sapien
Q9PKJ7 chlamydia m
Q9BDS2 lama glama
Q867B5 sus scrofa
Q9C8G1 arabidopsis
Q38971 arabidopsis
Q38970 arabidopsis
Q9SKV1 arabidopsis
Q7TRM0 mus musculu
Q9LDM4 oryza sativ
Q9UI31 leishmania
Q7X612 oryza sativ
Q37168 hepatitis g
Q9VSK2 drosophila
Q20219 caenorhabdi
Q9F6S9 streptomyce
Q80V41 mus musculu
Q95P13 strongyloce
Q94247 caenorhabdi
Q94246 caenorhabdi
Q9BEZ9 ursus arcto
Q90XK9 acipenser s
Q90XK8 poliodon sp
Q89XK8 bradyrhizob
Q802X1 brachydanio
Q92X67 rhizobium m
Q9VED3 drosophila
Q8W6J4 sinorhizobi
Q51556 pseudomonas
Q8GHA8 streptomyce
Q9Q6Q7 avian infec
Q60067 schizosacch

```

DR EMBL; AB008036; AAK6874.1; -.
DR PIR; A97490; A97490.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0016988; P:cell wall catabolism; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR005829; Sug_transporter.
DR Pfam; PF01476; LysM; 1.
DR SMART; SMO0257; LysM; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
SQ SEQUENCE 674 AA; 69487 MW; 5134DA44818141BC CRC64;

Alignment Scores:
Pred. No.: 0.295 Length: 674
Score: 85.50 Matches: 27
Percent Similarity: 43.59% Conservative: 7
Best Local Similarity: 34.62% Mismatches: 25
Query Match: 12.39% Indels: 19
DB: 16 Gaps: 3

M64347 (1-385) x Q8USE6 (1-674)
QY 376 CCACATGCTGGGCCCCAAGAA-----GAG 350
Db 310 ProProAlaLeuGlnProGlnAlaThrProSerGlnProSerAlaAlaThrSerGlu 329
QY 349 ACCACCTGAGCCATGGCCCTGAGGCAAGGACAGTCCCGAGTCTAG----- 296
Db 330 ThrThrAlaAsnThrProValProSerSerGlyThrAlaAlaProThrAlaAlaGly 349
QY 295 -----GGCCAGTAACA-----GTACAGAACGACCACTGAATTCACGGCT 254
Db 350 SerThrAlaValAlaProLeuThrSerSerValGlnValThrAlaValGluPheGluGly 369
QY 253 TCCCTCCCAAGTTTGAAGGTAGCAGTCCAGGCTATATAACTCTAGAACATTG 200
Db 370 SerLysIlePheValAlaGlySerAlaProGlyGlySerThrValArgAlaLeu 387

RESULT 2
Q96T16 PRELIMINARY; PRT; 504 AA.
ID Q96T16 AC Q96T16;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ14504.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Hago Y., Bahn A., Burchhardt G.;
RT "Molecular cloning and characterization of human sodium dicarboxylate
cotransporter 2 (hNADC2)".
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF489980; CAD34590.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006814; P:sodium ion transport; IEA.
DR InterPro; IPR001898; Na/sul symport.
DR Pfam; PF00939; Na sulph sympt; 1.
SQ SEQUENCE 440 AA; 47630 MW; DF0CDD54D9D1CC2 CRC64;

Alignment Scores:
Pred. No.: 1.14 Length: 440
Score: 80.50 Matches: 23
Percent Similarity: 47.83% Conservative: 10
Best Local Similarity: 33.33% Mismatches: 27
Query Match: 11.67% Indels: 9
DB: 4 Gaps: 2

M64347 (1-385) x Q8IV67 (1-440)
QY 385 CACCTCTGGCCACCATGCTGGCCCAAGAGAGACACCTGAGCCATGGCCCTGCA 326
Db 318 HisLeu-----CysGlyHisProAlaGlnSerIleLeuSerProSerProAla 334
QY 325 GCGAAGCAAGGACAGTGGCCCGAGTCCAGGCGCCAGTACAGTACAGAACCAACT 266
Db 335 GlyLeuGluGlyAsnProGlyGluSerAlaGlyAlaValArgValAspGlyGlu----- 352
QY 265 GAATTCAGGCTTCCCTCCCAAGCTTTGAAAGGTAGTCCAGGCTATAAATCTAGAA 206
Db 353 -----AlaHisGlyAlaLeuAlaArgSerAlaProGlySerHisProLeuAsp 368
QY 205 GCATTGGTGAAGAGTGTAAAGTCTAC 179
Db 369 PheValLeuAlaArgCysArgValHis 377

```

```

RESULT 4
Q86W50
ID Q86W50 PRELIMINARY; PRT; 562 AA.
AC Q86W50;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Hmo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; BC050503; AAH50603.1; -
KW Hypothetical protein.
SQ SEQUENCE 562 AA; 63648 MW; 003B5E5E4551F573F CRC64;

Alignment Scores:
Pred. No.: 1.34 Length: 562
Score: 80.00 Matches: 26
Percent Similarity: 45.71% Conservative: 6
Best Local Similarity: 37.14% Mismatches: 20
Query Match: 11.59% Indels: 18
DB: 4 Gaps: 3

M64347 (1-385) x Q86W50 (1-562)
QY 364 GGGCCCCAGAGAGACCCCTGAGCCATGGCCCTGAGGCAAGGACAGGACAGTGC 305
DB 425 GlyProGlnGluArgThrProCys-----GlyProAlaLeuArgGluGlyGluAla 442
QY 304 CAGACTCAGGGCCAGTACAGTACAGAAC-----GAACCA 269
DB 443 AlaValGluGlyProCysProSerGlnGluSerLeuSerGlnGluGluAsnProGluPro 462
QY 268 ACTGAATTCAGCGCTTCC-----CTCAAGCTTTGAAGCT 233
DB 463 ThrGluAspGluArgSerGluGluLysGlyValGluValLeuGluAsnCysGlnGly 482
QY 232 AGCAGTCCAGGCTATAAACTCTAGAAGCA 203
DB 483 SerSerAsnGlyAlaGlnAspGlnGluAla 492

RESULT 5
Q90Z54
ID Q90Z54 PRELIMINARY; PRT; 263 AA.
AC Q90Z54;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Caudal-type homeobox protein CDXB.
GN CDXB.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21391686; PubMed=11500978;
RA Ehrman L.A., Yutzy K.B.;
RT "Anterior expression of the caudal homologue cCdx-B activates a
RL Dev. Dyn. 221:412-421(2001).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR ENBL; AF353624; AAK38602.1; -
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.

GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR006820; Caudal_act.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF04731; Caudal_act; 1.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PF00024; HOMEBOX.
DR PRINTS; PF00031; HHREPRESSR.
DR PRODOM; PDC00010; Homeobox; 1.
DR SMART; SMO0389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 263 AA; 28817 MW; E87B71EBA1A7A695 CRC64;

Alignment Scores:
Pred. No.: 3.32 Length: 263
Score: 76.50 Matches: 38
Percent Similarity: 37.50% Conservative: 13
Best Local Similarity: 27.94% Mismatches: 53
Query Match: 11.09% Indels: 32
DB: 13 Gaps: 6

M64347 (1-385) x Q90Z54 (1-263)
QY 379 TGGCCACCATGCACCTGGGCCCCAAGAGAGACACCCCTGAGCCATGGCCCTGCA---GGC 323
DB 58 TrpGlyProHisTyrGlyProGlnArgGluAspTrpAsnAlaTyrGlyProCysProGly 77
QY 322 AAGCAAGGACAGTCCCGACTCAG-----GGCCAGTAACAGTACAGAAC 275
DB 78 ProSerGlyThrAlaAlaAlaGlnLeuSerGlySerProGlyGlnGlySerTyr 97
QY 274 GAACCACTGAATTCACGGCTTCCCTCCAGCTTTGAAAGGTAGCAGTCCA----- 224
DB 98 SerProAlaGluTyrSerSerLeuHisProAlaAlaProGlyGlyLeuProAlaGly 117
QY 223 -----GGCTATAAACTCTAGAACATTG 200
DB 118 ThrValGlyThrGlnGlnValSerProThrAsnGlnArgHisSerSerTyrGluTrpMet 137
QY 199 CGTAAGAGTGTAAAGTCTACACAAATACATCTGTATAAACTCAATAAATTATATA 140
DB 138 ArgLys-ThrVal-----ProThrAsnThrThrGlyLysThrArgThrLysGluLysTyr 155
QY 139 TAGATATATATAAACTGTAACTATATAACATCGGAACCTGCACACAGGCGCGCC 80
DB 155 rArgValValTyrThrAsp-----HisGlnArgLeuGluLeuGlu----- 168
QY 79 TCCCTGGAAACCGTCTCCCTCGCTGGGACACACAGCANTTAGACA 34
DB 169 -----LysGluPheHisCysAsnArgTyrIleThrIleArg 181

RESULT 6
Q17806
ID Q17806 PRELIMINARY; PRT; 263 AA.
AC Q17806;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE C08A9.3 protein.
GN C08A9.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150118; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
```

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMuray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Latreille P.;
 RT "The sequence of C. elegans cosmid C08A9";
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U42844; AAB53817.1; --
 DR PIR; TJ5453; TJ5453
 DR WORMPEP; C08A9.3; C506785.
 SQ SEQUENCE 263 AA; 30460 MW; EDF6EA44080AD135 CRC64;

Alignment Scores:
 Pred. No.: 5.78 Length: 263
 Score: 74.50 Matches: 29
 Percent Similarity: 39.29% Conservative: 15
 Best Local Similarity: 25.89% Mismatches: 25
 Query Match: 10.61% Indels: 43
 DB: 5 Gaps: 5

M64347 (1-385) x Q17806 (1-263)
 QY 28 ACAATTTCTTATTGCTGTGTCCTCCAGCAGGAGACGGTTCCAGGAGGGCCCG 87
 Db 140 SerAsnSerAlaIleCysLeu-----GlyArgGlyGln 151
 QY 88 CCCTGTGTGCAGGTTCCGATGTTATTAGATGTTACAGTTTATATATATATATATA 147
 Db 152 -----IleGlyIleTyrLeu 156
 QY 148 ATTATTGAGTTTACAGATGT----- 171
 Db 157 ValTyrTyrAlaValGlnLysCysArgPheGluArgGlnSerPheThrLeuPheTyrLys 176
 QY 172 ATTTGTTGTAGACATTAACTCTTACGCAATGTTCTAGAGTTTATATACCTGGACTGC 231
 Db 177 IleCysCysThrLeuIlePheIleValPheMetLeuMetGlu-----CysLeuAsnArg 194
 QY 232 TACCTTTCAAAGCTTGGAGGAGCCCGTGAATTCAGTTGTTCTGTTCTGTTACTGTACTG 291
 Db 195 TyrLeuAlaAsn-----PheMetLeuThrTyrAsnValLeuLeuLeuThr 208
 QY 292 GGCCCTGAGTCTGGCAGCTGCTCCCTGCTGCTGCTG 327
 Db 209 ProAlaLysSerArgGlnLeuAsnMetAspCysVal 220

RESULT 7

Q96T27
 ID Q96T27 PRELIMINARY; PRT; 867 AA.
 AC Q96T27;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Conserved hypothetical protein.
 GN B6L3.060.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 NCBI_TaxID=5141;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Algn V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Meves H.W., Mannhaupt G.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AU513462; CAD11405.1; --
 DR InterPro; IPR000433; Znf_ZZ.
 DR Pfam; PF00569; ZF_ZZ_2;
 DR SMART; SM00291; Znf_ZZ_4;
 DR PROSITE; PS0135; ZF_ZZ_2; 3.
 KW Hypothetical protein_
 SQ SEQUENCE 867 AA; 95865 MW; CE267925A35FE211 CRC64;

Alignment Scores:
 Pred. No.: 9.78 Length: 867
 Score: 73.00 Matches: 20
 Percent Similarity: 53.23% Conservative: 13
 Best Local Similarity: 32.26% Mismatches: 21
 Query Match: 10.58% Indels: 8
 DB: 3 Gaps: 2

M64347 (1-385) x Q96T27 (1-867)
 QY 361 CCCCAAGAGAGACCACTGAGCATGGCTGAGCAAGGAGGACAGCTGCCAG 302
 Db 110 ProGlnAspThrThrGluAlaLys---ProAlaSerValGluGlyGluAlaThrGlu 128
 QY 301 ACTCAGGCGCCAGTAAACA-----GTACAGAACGAAACCAACTGAA 263
 Db 129 ThrGluGlyAspAlaSerProSerAsnProThrSerProValTyrAspGluThrGlu 148
 QY 262 TTCAGCGCTTCCCTCCCAAGCTTTGAAGGTAGCTCCAGCTATAAAGCTAGAGCA 203
 Db 149 ValThrLysSerIleGluThrAspAlaSerProSerLysThrPheGluSerLeuGlnAla 168
 QY 202 TTGCGT 197
 Db 169 GlnArg 170

RESULT 8

Q9FQ80
 ID Q9FQ80 PRELIMINARY; PRT; 955 AA.
 AC Q9FQ80;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPCase).
 GN PPC.
 OS Chloris gayana.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Chloridoideae; Chlorideae; Chloris.
 NCBI_TaxID=110876;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fully expanded leaves;
 RA Blaesing O.E., Gowik U., Wolf B., Westhoff P.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: TO FORM OXALOACETATE, A FOUR-CARBON DICARBOXYLIC ACID
 CC -!- SOURCE FOR THE TRICARBOXYLIC ACID CYCLE (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: PHOSPHATE + OXALOACETATE = H(2)O +
 CC PHOSPHOENOLPYRUVATE + CO(2).
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -!- SIMILARITY: BELONGS TO THE PEPCase FAMILY.
 CC EMBL; AF268091; AAG42288.1; --
 DR HSP; P00864; 1PIY.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0008964; P:phosphoenolpyruvate carboxylase activity; IEA.
 DR GO; GO:0015977; P:carbon utilization by fixation of carbon di...; IEA.

M64347 (1-385) x Q18613 (1-474)

```
QY      11  CAAGCTGGTATTTCATACAAATTCTTCTAATTTGCTGTG-----TGTCCAGGCAG 61
Db      341  GlnSerGlyMetIleIleSerPheLeuIleLeuAlaIlePheLeuTrpIleAspTyrHis 360
QY      62  GGAGACGGTTTCCAGGAGGGCGCCCTGTGTGCAGGTCCGATGTTTATTAGATGTTA 121
Db      361  GlyThrGlyPhe-----Leu 365
QY      122  CAAAGTTTATATATCTATATATATATAATTTATTCAAGTTTTTACAAGATGTTATTGTTGA 181
Db      366  IleValTyrLeuPheGlyThrMet--PheIleGluTyrThrTrpAspAla----- 381
QY      182  GACTTAACACTTCTTACCGAATGCTTACAGATTATACCGCTGGAGCTGCTACCTTTCAA 241
Db      382  -----ValTyrLeuCysAlaIleGlu 388
QY      242  AGCTTGGAGGGAAGCGGTGAATTCAAGTTGCTGTTCTGTTACTGTTACTGGGCCCTGAGT 301
Db      389  SerMetGluThrSerSerArg-----AlaSerAla 398
QY      302  CTGGGCAAGCTGTCCCTTCTGCTGCCTGCAAGGCCCAAGCTCAGGCTGGTCTCTTCTTGGGG 361
Db      399  ValGlySerCysSerLeu-----MetAlaArgIleGlySerLeuLeuAla 413
QY      362  CCC-----AGTGCATGGTGGCCA 379
Db      414  ProPheLeuThrTyrAlaAsnThrTrpTrp 424

RESULT 10
Q97UC7 PRELIMINARY; PRT; 324 AA.
AC Q97UC7;
AD Q97UC7;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein SSO3088.
GN SSO3088.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=2132296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awaye M.J., Chan-Weiher C.C.-X., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fleischer C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.,
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006899; AAK43194.1; -.
DR PIR; C90492; C90492.
DR Interpro; IPR001455; UPF0033.
DR PROSITE; PS01148; UPF0033; 1.
KW Hypothetical protein; Complete proteome.
SO SEQUENCE 324 AA; 38040 MW; 861EF3865FDB448E CRC64;
```

QY 1 GACTTCAAGCAAGCTGTTATTTTCATACAAATCTCTCTAATTGCTGTGTGTCAGGCA 60
 Db 253 AspPheLys---AsnTyrArgPheHisCysGluLysAspIleCysLeuValSerSerAla 271
 QY 61 GGGAGACGGTTTCCA---GGGAGGGCCGGCCCTGTGTGTCAGGTTCCGATGTTATTAGAT 117
 Db 272 -----ProLeuGlyArgGlyTyr----- 277
 QY 118 GTTACAAGTTTATATATATCTATATATAATTATTAGTTTATACA----- 165
 Db 278 -----LeuPheThrGlyLeuLeuLys 284
 QY 166 -----AGAATGATTGTTGTAGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTAT 219
 Db 285 SerAsnLysMetValCysAlaArgIleAspThrGluAenGluThrLeuLeuAspTyrGln 304
 QY 220 AGCCTGGACTGTACTCTTTCARAGCTTGGAGGGAAG 255
 Db 305 AlaLeuGlu---TyrLeuLysLysLeuAlaGlyGlu 315
 RESULT 11
 ID O04114 PRELIMINARY; PRT; 447 AA.
 AC O04114;
 DT 01-JUL-1997 (TremBLrel. 04, Created)
 DT 01-JUL-1997 (TremBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
 DE Flavonoid 3-O-glucosyltransferase.
 OS Perilla frutescens.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; asterids;
 OC Lamiales; Lamiaceae; Nepetoideae; Elsholtziaceae; Perilla.
 OX NCBI_TaxID=48386;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98088016; PubMed=9426610;
 RA Gong Z., Yamazaki M., Sugiyama M., Tanaka Y., Saito K.,
 RT "Cloning and molecular analysis of structural genes involved in
 RT anthocyanin biosynthesis and expressed in a forma-specific manner in
 RT Perilla frutescens."
 RL Plant Mol. Biol. 35:915-927(1997).
 DR EMBL; AB002818; BAAL19659.1; -.
 DR GO; GO:0016740; F:transferase activity, transferring hexosyl . . . ; IEA.
 DR GO; GO:0006152; P:metabolism; IEA.
 DR InterPro; IPR002213; UDP-glucose trans.
 DR Pfam; PF00201; UDPGT; 1.
 DR PROSITE; PS00375; UDPGT; 1.
 KW Transferase.
 SQ SEQUENCE 447 AA; 48884 MW; 79B4C740FEB8ACBF CRC64;
 Alignment Scores:
 Pred. No.: 16.1 Length: 447
 Score: 71.00 Matches: 30
 Percent Similarity: 39.82% Conservative: 15
 Best Local Similarity: 26.55% Mismatches: 54
 Query Match: 10.11% Indels: 14
 DB: 10 Gaps: 4
 M64347 (1-385) x O04114 (1-447)
 QY 2 ACTTCAAGCAAGCTGTTATTTTCATACAAATCTCTAATTGCTGTGTGTCAGGCAAG 61
 Db 259 ThrSerProLysSerValValTyrIleSerPheGlyThrValIleThrProGluAen 278
 QY 62 GGAGACGGTTCCAGGAGGGCCGGCCCTGTGTGTCAGGTTCCGATGTTATTAGATGTA 121
 Db 279 GluLeuAlaLeuAlaAspAlaLeuGluIleCysArgPheProPheLeuTyrSerLeu 298
 QY 122 CAAGTTTAT 181
 Db 299 LysAspTyrAla-----ValLysSerLeuProAspGlyPheLeuAsp 312
 RP SEQUENCE FROM N.A.
 RX Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

QY 182 GACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGACTGCTACTTTCOA 241
 Db 313 -----ArgThrLysGlyPheGlyLysIleValAlaTyr---AlaProGlnGln 327
 QY 242 AGCTTGGAGGAAGCGCTGAATTCAGTTGCTTCTGTTCTGTTACTGTTACTGG----- 292
 Db 328 GlnValLeuAlaHisArgAsnValGlyValPheValThrHisCysGlyTyrPasnSerIle 347
 QY 293 GCCTGAGTCTGGCAGCTGTCCCTTGTCTGCTGCAGG 331
 Db 348 LeuGluSerIleSerCysValProLeuIleCysArg 360
 RESULT 12
 ID O81Q18 PRELIMINARY; PRT; 2114 AA.
 AC O81Q18;
 DT 01-MAR-2003 (TremBLrel. 23, Created)
 DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
 DE CG31968-PA.
 GN DP OR CG15637 OR CG31724 OR CG31725 OR CG31726 OR CG31727 OR CG31964
 OR CG31965 OR CG31966 OR CG31967 OR CG31968 OR CG33000 OR CG33086 OR
 GN CG33196.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.D.,
 RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Fabios B., Delecher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jallali B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy J., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Spadling A.C., Stalton M., Skupski M.P., Smith T.,
 RA Spier E.C., Strickland A.C., Stupple M., Strong R., Sun E.,
 RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Ye J.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou Q., Zhu L.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

M64347 (1-385) x Q9PKJ7 (1-360)

QY 91 TGTTGTCAGGTCGAGTATTATAGATGTTACAAAGTTATAT- 132
 Db 178 CysileLeuLeuProIleilethrValValArgValLeuThrAsnAlaPheArgPheLeu 197
 QY 133 ---ATATCTATATATATATTTAGTTTTCACAGATGT- 171
 Db 198 LeulleProPheTyrlleValPheGlnMetIleArgGlnLeuTyrglnGluAspLeuPro 217
 QY 172 -----ATTGTGTGT-----AGACTTAACACTTCTTAC----- 198
 Db 218 PheGluGluGlnPheIleCysSerAspIlePheArgGluMetSerArgSerPheValGln 237
 QY 199 GCAATGCTTCTAGAGTTTATAGCTGACCTGACCTACTCTTCAAGCTTGGAGGAGCCG 258
 Db 238 AlaValLysAlaProPheTyrglyValAlaCysTyrlleAlaSerLeuTyrgly- 255
 QY 259 TGAATTCAGTTGGTTGCTTCTGTTACTGTTACTTGGCCCTGAGTCTGGGCAGCTGTCCCTT 318
 Db 256 -----LeuLeuAsnProLeuSerGlyArgValIleMet 266
 QY 319 GCT----- 321
 Db 267 AlaSerValGluArgAspTrpAsnAsnAspValIleArgSerArgGlyIleTrpGlyIle 286
 QY 322 -----TGCTCGAGGCCATGCTCA----- 342
 Db 287 PheCysGluLysAsnCysLeuLeuGluGlyGlyThrArgSerGlyLeuGlyGlnHis 306
 QY 343 GGGTGGTCTCTTCTTGGG 360
 Db 307 AlaTrpTyrlleLeuGly 312

RESULT 15

Q7XNX1 PRELIMINARY; PRT; 735 AA.
 ID Q7XNX1
 AC Q7XNX1;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE OSJNB0026112.9 protein.
 GN OSJNB0026112.9
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
 RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
 RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
 RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Yang K., Zhou B., Chen Z.H.,
 RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
 RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
 RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
 RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
 RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL663002; CAB03901.1; -;
 SQ SEQUENCE 735 AA; 82524 MW; 4EF30D8946AC912A CRC64;

Alignment Scores:

Pred. No.:	19.3	Length:	735
Score:	70.50	Matches:	25
Percent Similarity:	43.18%	Conservative:	13
Best Local Similarity:	28.41%	Mismatches:	30
Query Match:	10.04%	Indels:	20
DB:	10	Gaps:	3

M64347 (1-385) x Q7XNX1 (1-735)

QY 159 TTTTACAAGATGATTGTTGTAGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTA 218
 Db 14 PheTyrglyAspPheAlaLeuProThrThrLysPhePheArggly-IleleuGluPheTy 33
 QY 219 TAGCCTCGACTGCTAC----- 234
 Db 33 rGlyIleAsnIleTyHisLeuAsnProAsnSerIleValHisIleAlaAsnPheValHi 53
 QY 235 -----CTTTCAAAGCTTGGAGGGAAGCCGTAATTCAGTTGGTTGCTTCTGTA 287
 Db 53 sLeuPheGluAlaPheLeuGlyIleArgProHisPheAlaLeuIleArgArgIlePhePh 73
 QY 288 ACTGGGCCCTGAGCTGGGCAGCTGCCCTTCCCTTCCCTGCGAGGCCCATGGTCTCAGGGTG 347
 Db 73 eLeuLysProGlnProAsnLys---AsnLysProCysIleValGly---GlySerSerPh 91
 QY 348 GTCTCTTCTTGGGGCCCACTGC 369
 Db 91 eGlnLeuArgGlyThrLeuCys 98

Search completed: September 22, 2004, 10:54:33

Job time : 98.5 secs